

A novel approach for predicting usability of upper limb prostheses

Junho Park^a, Maryam Zahabi^{b,*}, He Huang^c, Mark Benden^d

^a Department of General Engineering, Santa Clara University, Santa Clara, CA, USA

^b Wm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, College Station, TX, USA

^c Joint Dept. of Biomedical Engineering, North Carolina State University and University of North Carolina at Chapel Hill, North Carolina, USA

^d Department of Environmental & Occupational Health, School of Public Health, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Keywords:
Usability
Prosthesis
Upper limb
Prediction
Human performance model

ABSTRACT

Limb amputation can lead to significant functional challenges in daily activities, prompting amputees to use prosthetic devices (PDs). However, the cognitive demands of PDs and usability issues have resulted in user rejections. This study aimed to create a Human Performance Model for Upper-Limb Prosthetic Devices (HPM-UP). The model used formulations of learnability, error rate, memory load, efficiency, and satisfaction to assess usability. The model was validated in an experiment with 30 healthy participants using a bypass prosthetic device. Findings indicated that the HPM-UP successfully predicted the usability of prosthetic devices, aligning with human subject data. This research proposes a quantitative approach to predict upper limb prosthetic device usability by quantifying each dimension and computationally connecting them. The model, available on Github and executable with Rstudio, could enable clinicians to assess and analyze the human performance of various commercial prostheses, aiding in recommending optimal devices for patients.

1. Introduction

The lack of useable prosthetic devices causes amputee patients to have difficulties in performing activities of daily living (ADLs). In the U.S., there are over two million people who have lost a limb, and this number is projected to increase twofold by 2050 (Niamba et al., 2021). Prosthetic devices are essential for ADLs, but they may also demand more time and effort from the users (Park et al., 2022; Park and Zahabi, 2022a). Many users find the existing devices hard to operate, resulting in lower usage and device abandonment (Engdahl et al., 2015). A study that evaluated the usability of various prosthetic devices reported that 53% of passive hand users, 50% of body-powered hook users, and 39% of myoelectric hand users gave up on using prosthetic arms (Montagnani et al., 2015) mainly because of poor dexterity, glove durability, and lack of sensory feedback.

The International Standard Organization (ISO) defines usability as “The extent to which specified users can use a product to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.” (ISO, 2019). Usability is a critical factor in the adoption and continued use of prosthetic devices. Prosthetic device rejection, where users decide not to use or abandon their prosthetic limbs, can often be

attributed to issues related to the usability of these devices. The primary aspects of usability that influence prosthetic acceptance include comfort, functionality, and the ease with which a user can operate the device. Devices that cause discomfort, pain, or do not fit properly are likely to be rejected. Prior studies found that discomfort and poor fit are major reasons for the abandonment of prosthetic devices (Biddiss and Chau, 2007). Functionality refers to how well the prosthetic meets the user's needs in terms of movement and daily activities. A prosthetic that does not adequately replace the functionality of a lost limb or fails under specific conditions is less likely to be accepted by the user. Functional limitations have been cited as significant factors influencing the rejection of prosthetic devices (Sanders and Fatone, 2011). The ease of use of a prosthetic device also plays a crucial role. Devices that are complex to operate or require extensive maintenance can be frustrating for users, leading to abandonment. Enhancing usability by focusing on the ease of use can significantly improve acceptance rates (Peerdeman et al., 2011). Addressing these usability issues through user-centered design and regular feedback sessions can help mitigate the risk of prosthetic device rejection.

Nielsen (2012) proposed five dimensions of usability, including: Learnability, efficiency, memorability, error, and satisfaction.

* Corresponding author. Wm Michael Barnes '64 Department of Industrial & Systems Engineering, Texas A&M University, Emerging Technologies Building, College Station, TX, 77843, USA

E-mail address: mzahabi@tamu.edu (M. Zahabi).

<https://doi.org/10.1016/j.apergo.2024.104344>

Received 14 January 2024; Received in revised form 23 June 2024; Accepted 2 July 2024

Available online 10 July 2024

0003-6870/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Learnability is an indicator that shows how users reach optimal performance in interacting with a system (Joyce, 2019). Alternatively, it can be interpreted as how easy it is for users to accomplish a task the first time they encounter the interface and how many repetitions it takes to become efficient at that task. Learnability can be measured using learning time or the number of trials in a training session. *Efficiency* refers to how fast users can perform tasks once they have learned to use the system and is typically measured using the task completion time (TCT) (Dix et al., 2000). *Memorability* means the system should be easy to remember so that the users can return to the system after some time without learning everything all over again. An *error* can be defined as a function performed by a user that does not lead to the aimed result. Finally, *satisfaction* refers to how pleasant the system is to use. Also, it refers to the level of comfort and acceptability of the system to its users and other people affected by its use (Dix et al., 2000).

Usability of prosthetic devices has been measured by several subjective questionnaires, such as the Client Satisfaction with Device module of the Orthotics and Prosthetic Users' Survey (CSD-OPUS) (Bravini et al., 2014). This is a self-report instrument for evaluating the outcomes (satisfaction) of prosthetics and orthotics. Another questionnaire used in this domain is Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0) (Demers et al., 2002). This questionnaire is designed for a person's evaluation of those distinct dimensions of the assistive device that are influenced by one's expectations, perceptions, attitudes, and personal values. The third questionnaire is Usefulness, Satisfaction, and Ease of Use (USE) survey (Lund, 2001), which measures subjective usability and has been applied for evaluating prosthetic devices and other products (e.g., wearable devices, smartphones, website). Finally, the System Usability Scale (SUS), which is also designed for products or services (Brooke, 1996), has been used to assess the usability of prosthetic devices. SUS evaluates various products and services, including hardware, software, mobile devices, and websites. SUS and USE surveys are more focused on subjective attributes such as comfort, ease of use, satisfaction, and willingness to use as they are heavily used in assessing the usability of websites or software (Gao et al., 2018; McLellan et al., 2012). Meanwhile, the CSD-OPUS and QUEST 2.0 surveys incorporate physical attributes such as weight or aesthetic aspects of prosthetic devices.

Although there are several questionnaires for assessing the usability of prostheses, these methods are mainly used toward the later stages of the design and development process when there is a functional prosthetic device and there is a need for conducting a human subject study which can be costly and time consuming. Furthermore, the scales provide discrete rather than continuous values, and prior studies have found a dissociation between subjective usability evaluation and objective performance measures (Yeh and Wickens, 1988). Also, there could be self-report biases with survey responses (Hart, 2006).

To fill these research gaps, this study aimed to develop a usability prediction model for prosthetic devices (called HPM-UP). The main motivation for the model development is to have an analytical approach to overcome the limitations of subjective evaluation techniques in prosthetic assessments and to predict usability of prostheses in early stages of the design process.

2. Method

2.1. Human subject experiment

2.1.1. Participants

Thirty healthy participants (18 males and 12 females) were recruited for this study (Age: $M = 22.9$ yrs; $SD = 2.8$ yrs). All participants had 20/20 vision (or corrected vision) and no prior experience of using prostheses or myoelectric exoskeleton for upper limbs. The experiment protocol was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill.

2.1.2. Apparatus

A commercial 2-degrees of freedom (DoF) prosthetic device (Motion Control ETD, Filauer) in hand open/close and wrist pronation/supination was used in this study. A custom prosthetic hand adapter was designed and fabricated as a bypass device, as shown in Fig. 1. This experiment included testing of three prosthetic device configurations: (1) direct control (DC), (2) pattern recognition (PR), and (3) continuous control (CC). For the DC mode, EMG signals were collected from two channels (the flexor carpi radialis and the extensor carpi radialis longus) based on the mean absolute value (MAV) of each channel (Resnik et al., 2018; White et al., 2017). Participants were trained with five hand gestures (hand close/open, wrist pronation/supination, inactive) in the PR mode. EMG data were collected and labeled simultaneously with a certain movement class. Four commonly used time domain features (MAV, number of zero crossings, waveform length, and number of slope sign changes) were extracted from EMG signals following the methods used in prior studies (Resnik et al., 2018; White et al., 2017). In the CC mode, EMG data were recorded simultaneously with kinematic data from a Leap Motion Controller (Leap Motion, Inc., USA). An artificial neural network was created for each participant using the Deep Learning Toolbox in MATLAB 2018b (Mathworks Inc., USA). Hand gestures and hook movements for each control scheme is summarized in Table 1. More information about the apparatus and sensor placement can be found in Park et al. (2023) and Park (2023).

2.1.3. Experiment design

The experiment followed a between-subject design, in which each participant was randomly assigned to one of the three prosthetic configurations (i.e., DC, PR, or CC) to avoid potential fatigue or learning effect from one trial to the next. Clothespin Relocation Test (CRT) and Southampton Hand Assessment Procedure (SHAP) were used as ADLs in this experiment as they have been widely used for assessing usability of upper limb prostheses (Zahabi et al., 2019). The CRT required participants to move as many pins as possible from the horizontal bar to the vertical bar and vice versa in 2 min with various hand gestures in each control mode (Park et al., 2020) (Fig. 2). For SHAP, participants had to turn the door handle down and bring it back five times as quickly as possible (Fig. 3).

2.1.4. Variables

This study compared the usability estimates from the Human Performance Model for Upper-limb Prostheses (HPM-UP) method, human-subject experiment, and a benchmark model. The benchmark model was developed using the Cognitive-Perceptual-Motor – Goals, Operators, Methods, and Selection rules (CPM-GOMS) method and Adaptive Control of Thought—Rational (ACT-R) working memory module in Cogulator software (Estes, 2017) to be compared with HPM-UP and human-subject experiment outcomes.

The dependent variables included task performance and usability measures (from QUEST 2.0 (Demers et al., 2002) and USE (Lund, 2001) questionnaires) and were used as a ground truth when comparing the outcomes with model estimates. Task performance measures included

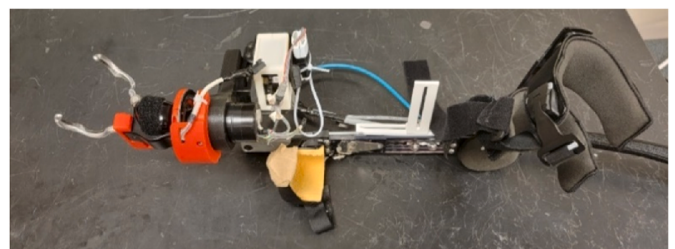


Fig. 1. The prosthetic device used in this study.

Table 1
Hand gestures and hook movements.

Prosthetic Movement	Pattern Recognition & Continuous Control Hand Movement	Direct Control Hand Movement
		
Open Prosthetic	Open Hand	Extend Hand at Wrist
		
Close Prosthetic Mode Change	Close Hand Not Applicable	Flex Hand at Wrist
		
Supinate (Rotate Clockwise) Prosthetic	Supinate Hand	Power Grip to Change Between Mode 1 and 2
		
Pronate (Rotate Counterclockwise) Prosthetic	Pronate Hand	Extend Hand at Wrist
		
		Flex Hand at Wrist

the number of pins moved within 2 min for the CRT and the time to rotate the door handle five times sequentially for the SHAP door handle task.

2.1.5. Procedure

Once participants arrived at the lab, they signed the informed consent form, and filled out the demographic questionnaire. Then, they completed the Edinburgh Handedness Test (EHT) (Oldfield, 1971) and the Purdue Pegboard Test (PPT) (Tiffin and Asher, 1948; White et al., 2017) to assess dexterity and right handedness. Once participants completed the EHT and PPT, they were equipped with the prosthesis and EMG electrodes were placed on their skin based on the assigned control mode. Participants were allowed to interact with the device and perform different movements (i.e., opening and closing the hand, wrist pronation, and supination) until they reported comfort with the control mode and the classifier was sufficiently trained if it could classify the gestures with accuracy of >90% (Fontana, 2010; Parajuli et al., 2019). Once the participants received training for their assigned control mode, they were trained on the CRT, which assessed the mastery of device handling and the respective control mode. If the average TCT to move three pins in three sequential trials was within 15–25s for the PR, 20–35s for the DC (White et al., 2017), and 16–23s for the CC mode trials (Park, 2023), the participant passed the training criteria and was allowed to proceed to the testing session. We used different TCT thresholds because of the differences in the DoFs, types of gestures, and control modes. The thresholds for the DC and PR modes were defined based on our previous

investigations using similar prosthetic devices (White et al., 2017) and the threshold for the CC mode was defined based on our pilot testing.

Upon completion of the training trials, participants could begin the experimental trials after having 5 min of rest. In experimental trials, participants were instructed to move as many clothespins as possible between the 2 bars within 2 min for the CRT task. For SHAP – Door Handle, participants were instructed to rotate the handle five times as fast as possible. All participants completed three trials for each task with a 5-min rest period after each trial. At the end of the experiment, participants filled out the USE and QUEST 2.0 questionnaires.

2.1.6. Hypotheses

Five hypotheses (H) were formulated for this study (Table 2). Since both human subject data and the benchmark model did not generate all usability dimensions, hypotheses were generated to enable comparisons between the HPM-UP and either human data or the benchmark model.

2.1.7. Data analysis

The statistical model did not meet the parametric assumptions and therefore, nonparametric analysis was conducted to assess the differences in usability dimensions among the human subject data, HPM-UP, and the benchmark model. For the comparison between two sets of data, Wilcoxon rank sum test was conducted (Wilcoxon, 1992). The Wilcoxon test statistic *W* was used to determine the significance of the difference. Kruskal-Wallis rank sum test was conducted if there were more than two groups of data (Kruskal and Wallis, 1952). *H* statistic was used to

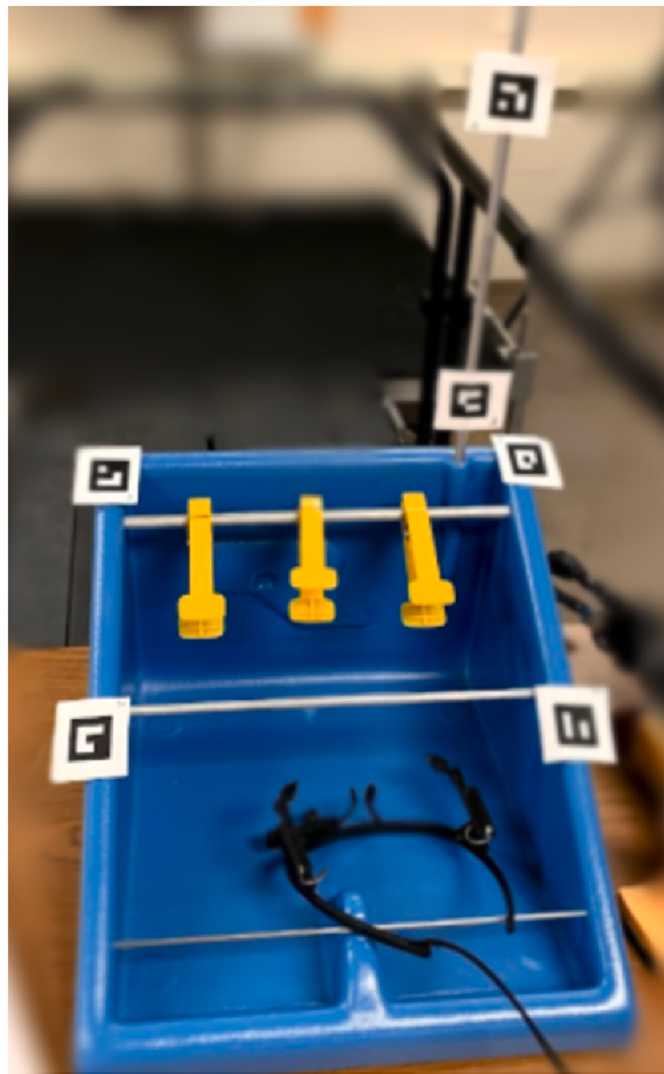


Fig. 2. The clothespin relocation task.

determine the significance of the difference of the median of each group. *H* statistics was compared with a critical cutoff point determined by the chi-square distribution (chi-square is used because it is a good approximation of *H*, especially if each group’s sample size is bigger than 5). For the post-hoc analysis, Dunn’s Kruskal-Wallis multiple comparison was used (Dunn, 1964). All the statistical analysis was conducted using R 4.0.5. Effect size for Wilcoxon signed-rank test was calculated with $r = \frac{Z}{\sqrt{n}}$ where *Z*-score is a test statistic and standardized score of *U*-value calculated from Mann-Whitney *U* test (Tomczak and Tomczak, 2014) and *n* is the total number of observations. The effect size of Kruskal-Wallis test was calculated using Eta-squared (Rosenthal, 1986).

2.2. Model development

HPM-UP was developed based on the theories in human factors and psychology, and human-subject data. The model stemmed from the CPM-GOMS method since some tasks can occur in parallel (John and Gray, 1994) and thus, CPM-GOMS provide a more accurate TCT estimates as compared to other models that assume serial activities. In CPM-GOMS models, the task is broken down into a series of operators or actions. Each operator is assigned a specific execution time (e.g., keystroke = 0.28 s) (Kieras and Santoro, 2004). These times are typically determined by previous research or user testing. The next step involves constructing a schedule chart, similar to those used in project



Fig. 3. The SHAP door handle task.

Table 2
List of hypotheses.

Hypothesis ID	Hypothesis
H1 (Learnability)	The results of HPM-UP learnability dimension would be similar to the human-subject data.
H2 (Error rate)	Use of the CC configuration would lead to the lowest error rate followed by the PR and DC configurations.
H3 (Memory load)	The results of HPM-UP memory load dimension would be similar to the benchmark model.
H4 (Efficiency)	(H4-1) The results of HPM-UP efficiency dimension would be similar to the human subject data. (H4-2) There would be a significant difference between the HPM-UP efficiency dimension results and the benchmark model results.
H5 (Satisfaction)	The results of HPM-UP satisfaction dimension would be similar to the human-subject data.

management (John and Kieras, 1994). TCT is then calculated by adding up the execution times for all the operators. However, unlike other GOMS models, CPM-GOMS takes into account the possibility of parallelism. That is, it recognizes that some operators can be performed simultaneously, rather than sequentially. This allows CPM-GOMS to provide a more realistic estimate of TCT for expert users, who often overlap their activities to work more efficiently.

In addition to the CPM-GOMS logic, declarative memory module functions from ACT-R 7.0 were used for simulating the number of memory chunks in HPM-UP (Bothell, 2017; Estes, 2015; Leiden and Best, 2005). For example, while performing ADLs, participants had to remember a particular device configuration to adjust the hook and complete the tasks. R software package 4.0.5 was used for model development.

2.2.1. Outcomes of the model

HPM-UP provides five usability dimensions based on Nielsen (2012) five dimensions of usability, including learnability, efficiency, memory load, errors, and satisfaction as these dimensions have been frequently used in evaluating the usability of other products. For the HPM-UP, there was a need to have new definitions of usability dimensions related to the prosthetic device application. Therefore, the modified usability dimensions were defined as follows.

- **Learnability:** The number of training trials required to pass the training criteria
- **Error (Error Rate):** The error rate in performing a task with a prosthetic device
- **Memory Load:** The number of memory chunks stored in working memory when performing a task with a prosthetic device
- **Efficiency:** TCT of one cycle of an ADL task
- **Satisfaction:** The relationship among perceived performance, expectation, and desire (Oliver, 1977, 1980)

These dimensions are described in detail below.

2.2.2. Learnability

The learnability dimension in HPM-UP is defined as the number of training trials to pass the training threshold level (or to reach a plateau). The original learnability equation is based on the learning curve's unit theory (Camm, 1985; Mislick and Nussbaum, 2015; Zhang et al., 2016) which is defined as Equation (1).

$$Y_x = Ax^b \quad (1)$$

In equation (1), Y_x is the cost of unit x , A is the theoretical cost of unit 1, x is the unit number, and b is a constant that determines the slope of the learning curve. In HPM-UP, x is the number of training trials which is the outcome of learnability. A is the TCT for the first trial. Y_x can be replaced with the TCT in each trial and can be gathered from training trials.

Pilot data from 10 participants were used to investigate the patterns in TCT (Park, 2023). It was found that TCT of the first trial (i.e., A) is different across participants, which might be due to the device calibration quality (DCQ). There are two types of DCQ including: Objective DCQ (ODCQ) and subjective DCQ (SDCQ). ODCQ can be calculated using Equation (2) (Music, 2022).

$$ODCQ = \frac{\text{The number of correctly matched input gestures}}{\text{The number of all input gestures}} \quad (2)$$

ODCQ can be used to estimate A for each participant, however, it requires collecting data from all participants and conducting a detailed video analysis, which might not be suitable for the HPM-UP package as the goal of this package is to use it in early stages of the design and development process of prosthetic devices. Additionally, video analysis cannot guarantee if specific activities were intended or not by the participants. Therefore, the SDCQ was used to adjust A . Having high SDCQ could result in small A (i.e., positive perception of device calibration quality could lead to shorter TCT in the first trial). However, subjective evaluations can have *self-report bias* (VandenBos, 2007). To mitigate the self-report bias, the *first impression (FI)* of the prosthesis was also considered. *FI* refers to one's initial perception of a person (or object), typically involving a positive or negative evaluation as well as a sense of physical (or psychological) characteristics (VandenBos, 2007). In this study, although all the participants passed the training criteria, the number of required training trials were different among participants, which led to having different error rates, or efficiency in the experimental trials. These differences could have been due to different *FI* of the device.

Therefore, the A factor was adjusted (called A') based on the *FI* and SDCQ (Equation (3)). If SDCQ is 1, which means users think the DCQ is perfect or has no errors, the *FI* is the only concern. However, if SDCQ is

for example 0.5 (e.g., 5 out of 10 gesture inputs were correct), the A' will be doubled. The *FI* factor is a number between 0 and 1. If *FI* is less than 1, A' will increase. If *FI* equals to 1 (i.e., the *FI* is positive), A' will only be influenced by SDCQ.

$$A' = \frac{A}{SDCQ \cdot FI} \quad (3)$$

SDCQ in this study was calculated based on the average of the responses to questions Q3 (easiness in adjusting the device (fixing, fastening)), Q6 (easiness of using the device), and Q8 (effectiveness of using the device (the degree to which the device meets a user's needs) of the USE questionnaire. *FI* was calculated from the difference between the SDCQ and participant's training performance. This indirect approach was used to avoid self-report bias (VandenBos, 2007) that could have occurred if *FI* was measured directly. The slope of learning curve is represented by b as shown in Equation (4).

$$\text{slope of learning curve} = \frac{\text{cost of unit } 2n}{\text{cost of unit } n} = \frac{A(2n)^b}{A(n)^b} = 2^b \quad (4)$$

$$\ln(\text{slope}) = b \ln(2) \therefore b = \frac{\ln(\text{slope})}{\ln 2}$$

Previous studies found that the slope could be estimated based on the degree of automation (e.g., 70% for entirely manual operations) (Mislick and Nussbaum, 2015). Since the tasks in this study were performed by prostheses, ideally, the slope could be 0.7 when the calibration quality is perfect (Mislick and Nussbaum, 2015) (i.e., the prosthesis could always be controlled based on user's input). However, since calibration quality could be varied, there is a need to adjust the slope. If SDCQ becomes 0, the slope changes to 0.90 (i.e., there is almost no learning occurred between the trials - refers to the 25% manual operation) (Mislick and Nussbaum, 2015). Therefore, Equation (5) shows the linear relationship between the SDCQ and slope.

$$\text{slope} = -0.2(\text{SDCQ}) + 0.9 \quad (5)$$

Equation (5) does not include a potential effect of physical and/or mental workload on device learnability. Therefore, two parameters were added to adjust the slope based on the level of workload as shown Equation (6).

$$\text{slope} = -0.2R_1(\text{SDCQ}) + (0.9 + R_2) \quad (6)$$

The initial values for R_1 and R_2 parameters were determined based on the pilot study with 10 participants. R_2 refers to the baseline physical and mental demand of participants before the experiment which was quantified using the NASA-Task Load Index (TLX) questionnaire (Hart and Staveland, 1988). R_1 was determined based on the changes of physical and mental workload from the baseline.

Based on the changes in slope and A factor described earlier, the learning curve equation was revised to estimate learnability as shown in Equation (7).

$$L = A'x^b = \frac{A}{SDCQ \cdot FI} x^{\frac{\ln(-0.2R_1SDCQ + (0.9 + R_2))}{\ln 2}} \quad (7)$$

2.2.3. Error rate

Including errors in the HPM-UP for analysis of prosthetic devices is necessary since unlike the direct human interaction with a mouse or keyboard controlled with high precision, a prosthesis device uses an EMG signal that does not always correctly project user's intentions to the outcome (i.e., hook movement). That is, there is a mediator (prosthesis) between the human and the task, which can affect the task performance. Some prior studies estimated errors in using prostheses with the concepts of total and active errors (Hargrove et al., 2007; Lock et al., 2005; Scheme and Englehart, 2011). Active error is determined by the proportion of erroneous active decisions to the overall count of active

decisions (Music, 2022). Total error rate is computed as the ratio of incorrect decisions, specifically mismatches between the input gesture and hook movement, to the total number of decisions. However, these studies had some limitations. First, it is challenging to clearly figure out whether a specific gesture is an active or inactive decision from observational studies. Second, prior studies only considered the pattern recognition (PR) configuration, which might limit the generalizability of the approach to other prosthetic device configurations such as the direct control (DC) or continuous control (CC).

To address the identified limitations, the error rate in HPM-UP was formulated differently from Scheme and Englehart (2011) study. The error rate in HPM-UP depends on learnability (e.g., participants who reached the training criteria faster exhibited few numbers of errors in experimental trials). Error rate was estimated based on learnability (i.e., $Err(L)$) because of the causal relationship between two dimensions. If $L = 1$, (i.e., learnability is 100%), the estimated error rate in the experimental trials will be 0 (i.e., $Err(1) = 0$). If $L = 0$, $Err(0)$ will be 1, which means that participants will make errors during the experimental trials since they failed to learn how to use the device during the training. The error rate follows the natural exponential function as shown in Equation (8). This exponential curve was fitted based on the results of pilot testing using the calculated learnability and error rate observed during the experiment.

$$Err(L) = \max\left\{\frac{1}{1-e}(e^L - e), 0\right\} \quad (8)$$

2.2.4. Memory load

Declarative memories are the kind of memories that can be declared (e.g., the name of one's fifth grade math teacher). Chunks are the elements of declarative knowledge in the ACT-R theory and are used to communicate information among modules through the buffer (Bothell,

$$Disconfirmation\ of\ belief_i = Z \left\{ \bar{P}_i - f(\bar{L}_i, \bar{q}_i) \right\} = Z \left\{ Perceived\ performance_i - \frac{Entire\ task\ duration\ (i.e.,\ 120\ seconds)}{\min\ threshold_i + Learnability_i \times (\max\ threshold_i - \min\ threshold_i)} \times q_i \right\} \quad (11)$$

2020). Chunks help link past experiences with the present context, determining their relevance and ease of retrieval at any given time. HPM-UP measured each chunk's duration in working memory. Chunk's duration was measured as time in millisecond from the moment when it first entered the working memory stack and the moment when it disappeared from the working memory stack due to the time lapse without retrieval. Therefore, memory load in HPM-UP was defined as the overall occupancy or duration of working memory chunks in the entire task and was calculated using Equation (9), which divides the summation of the duration of all chunks by the total task duration (Estes, 2015, 2021). Thus, the memory load in HPM-UP is different from learnability, as it only considers the number of memory chunks and task duration.

$$Memory\ load = \frac{\sum_{i=1}^7 (chunk\ duration)_i}{total\ task\ duration} \quad (9)$$

2.2.5. Efficiency

Efficiency was calculated using Equation (10) and based on the TCT and considering error rate. HPM-UP calculates the efficiency of an expert (i.e., no errors) based on the TCT estimates from the CPM-GOMS method and inflates the task time based on the error rate to estimate efficiency for all users.

$$Eff(Err) = Eff\{Err(L)\} = (Expert's\ efficiency) \times \{1 + Err(L)\} \quad (10)$$

The time duration of general operators (e.g., "Grasp") follows the Motion-Time Measurement (MTM-1) (Maynard et al., 1948), as it has the most fine-grained level of description for human movements. More information regarding the use of these operators in our model can be found in Park (2023).

2.2.6. Satisfaction

The theoretical foundation to formulate *satisfaction* came from the expectation confirmation theory (ECT) which is a cognitive theory that explains satisfaction as a function of *expectations* and *perceived performance* (Oliver, 1977, 1980). Once users accumulate experience on a product or service, they can subjectively evaluate their performance with the device (Lowry et al., 2006). The user compares the *desire* and *expectations* against the perceived performance of the product. *Expectation* is a belief or subjective prediction about a product's attributes or performance at some point in the future (Bhattacharjee, 2001). *Perceived performance* is a user's perception of the degree to which a product can fulfill his or her expectation in actual usage. *Desire* is the level of attributes and benefits that leads to attaining the user's desired outcomes (Spreng et al., 1996). ECT posits that satisfaction is directly influenced by disconfirmation of beliefs and perceived performance and is indirectly influenced by both expectations and perceived performance by means of a mediational relationship which passes through the disconfirmation construct.

Expectation function (f) in the HPM-UP was defined based on the expected task performance after passing the training session (or expected performance before the experiment trials). *Desire* was determined with q that was used in the calculation of *learnability* dimension. *Perceived performance* was calculated from the *Efficiency* dimension. Based on these concepts, disconfirmation of beliefs was formulated as Equation (11).

In this equation, $\bar{P}_i$ is a matrix or vector of participants' perceived performance, which is calculated from the efficiency module of HPM-UP. *Expectation* (f) is a function of learnability ($\bar{L}_i$) and desire ($\bar{q}_i$) because *expectation* can be estimated using the training performance or *learnability* as it is a belief or subjective prediction of performance in the future. For example, if users pass the training criteria only within 3 or 4 trials, they may perform well during the experimental trials. Thus, *learnability* was used as a variable to determine *expectation* (i.e., expected performance) before the experimental trials based on the thresholds defined for each configuration ($threshold_i$; e.g., 20–35s for the DC configuration). Based on the definition of desire (i.e., the level of attributes and benefits that leads to attaining the user's desired outcomes), q ranges between 0 and 1 and is multiplied by *expectation*.

Lastly, *Effort* was defined as the level of difficulty (mentally and physically) in performing an activity (Hart and Staveland, 1988), which could also affect perceived performance. Therefore, the original ECT was revised to include the level of effort needed to perform the tasks (Fig. 4). Then, satisfaction was finally formulated with Equation (12) based on the disconfirmation of beliefs and effort. In addition, a constant value (c) was added which refers to the minimum level of satisfaction.

$$Satisfaction_i = (Disconfirmation\ of\ belief_i) \times \left(1 - \frac{Effort}{100}\right) + c \quad (12)$$

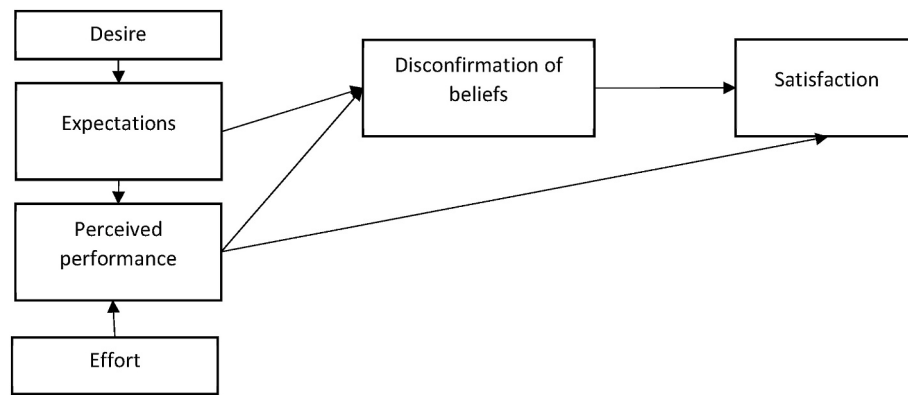


Fig. 4. Revised expectation-confirmation theory.

3. Results

3.1. HPM-UP

An overview of HPM-UP graphical user interface (GUI) is illustrated in Fig. 5. Analysts can start using HPM-UP by either using the ‘Develop a Scenario’ tab or loading an already developed scenario (Microsoft CSV format). Then, input parameters should be determined from the ‘Input parameters’ tab. Lastly, the model will assess the usability of the prosthetic device based on dimensions including learnability, error rate, memory load, efficiency, and satisfaction as shown in the ‘Results’ tab.

Figs. 6–8 illustrate the HPM-UP input entry, scenario development interface, and the output window respectively. In the input parameter screen (Fig. 6), the type of ADL task (e.g., CRT or SHAP) and device control scheme (e.g., DC, PR, or CC) can be selected. The minimum and maximum training time duration should be specified based on analysts’ previous knowledge, literature, or pilot test results (the minimum and maximum values in Fig. 6 came from our pilot test results based on the PR configuration). Device calibration quality (0–1), first impression (0–1), and effort (0–1) should be determined based on the analyst expertise and/or end users’ input.

If the analysts would like to develop a scenario manually, they can click the ‘Develop a Scenario’ tab (Fig. 7). Analysts can choose one of the appropriate operators or memory chunks from the radio buttons on the screen. If they would like to unselect radio buttons, they can click on the ‘Remove current selections’ button. First, analysts should define a goal for their model (by clicking the ‘Add a goal?’ check box). Once the goal is described in the text box of ‘Describe the use of the operator,’ the analysts can click ‘Add new line to code’ to add a line of code to the scenario. The analyst can continuously develop the scenario by clicking one of the operators and describing the operator in the related textbox. Parallel activities can be added by clicking the “Parallel?” check box when adding a line to the scenario. A chunk can be added to the scenario once analysts choose an operator, click one of the chunks under the “Chunks” column, describe the operator, and click “Add new line to Code”. Analysts can also add a custom chunk to the scenario. If there is a need to add new or customized operators, the analyst can choose “custom” at the bottom of each column of perceptual, cognitive, or motor operators. Then, the name of operator and duration can be specified. The scenario can also be downloaded in a CSV format once it is complete. This is useful because working or editing directly on a CSV format file might be necessary when the analysts are developing more complex scenarios. The CSV format file can also be loaded from the HPM-UP main screen.

Once all the input parameters are added, analysts can see the outcomes in terms of the predicted five usability dimensions (Fig. 8).

To provide a rule of thumb to support the interpretation of the outcomes of HPM-UP (Fig. 8), a literature review was conducted (Table 3). However, it is important to note that the outcomes depend on several

factors such as individuals’ physical condition, the amount of time they spend using the prosthesis per day, and the complexity of the tasks they are performing with it.

3.2. Model validation

Table 4 illustrates the descriptive statistics results regarding the comparison among the human subject data, the HPM-UP modeling approach, and the benchmark model. The benchmark model does not provide learnability, error rate, and satisfaction, and therefore, these cells are marked with not applicable (N/A). However, HPM-UP was able to generate all five dimensions.

A summary of hypothesis test results is shown in Table 5. All the hypotheses were supported except for H2. For the *learnability* dimension (H1), there was no significant difference between the human subject data and HPM-UP generated data based on the Wilcoxon Signed-Ranks Sum test ($W = 436.5, p > 0.05$).

Hypothesis 2 was refuted as there was no significant difference among configurations in terms of *error rate* ($H(2) = 1.57, p > 0.05$). According to the computational logic of error rate in HPM-UP, the error rate depends on learnability. There was no significant difference in *learnability* among different configurations from the human subject data, which led to not having any significant difference in error rate as well.

There was no significant difference in *memory load* between the HPM-UP and the benchmark model (H3) ($W = 65, p > 0.05$). In the DC configuration, participants needed to memorize and recall two mode changes (supination/pronation or open/close) and three gestures (open, close, AND rotation). Using the PR and CC configurations did not involve memory chunks as these configurations were more intuitive.

For efficiency, both hypotheses (i.e., H4-1 and H4-2) were supported. Based on Dunn’s Kruskal-Wallis multiple comparison, there was no significant difference between the human subject data and HPM-UP efficiency outcomes (H4-1) ($Z = 0.26, p > 0.05$). However, there was a significant difference between HPM-UP and the benchmark model (H4-2) ($Z = -4.54, p < 0.001$). There was also no significant difference in *satisfaction* between the human subject data (USE questionnaire – satisfaction dimension) and HPM-UP ($W=41, p > 0.05$).

To further validate the model, the root mean square error (RMSE) and *R-squared* were calculated (Wu, 2018). RMSE was used to measure the average deviation between the predicted value and the observed data, while *R-squared* was used to understand how well the predictor variables can explain the variation in the response variable. RMSE for learnability between the observation and HPM-UP was 2.1 with an *R-squared* value of 0.56. For efficiency, RMSE was 2.42s and *R-squared* was 0.48. The RMSE for satisfaction was 0.19 with *R-squared* value of 0.89.

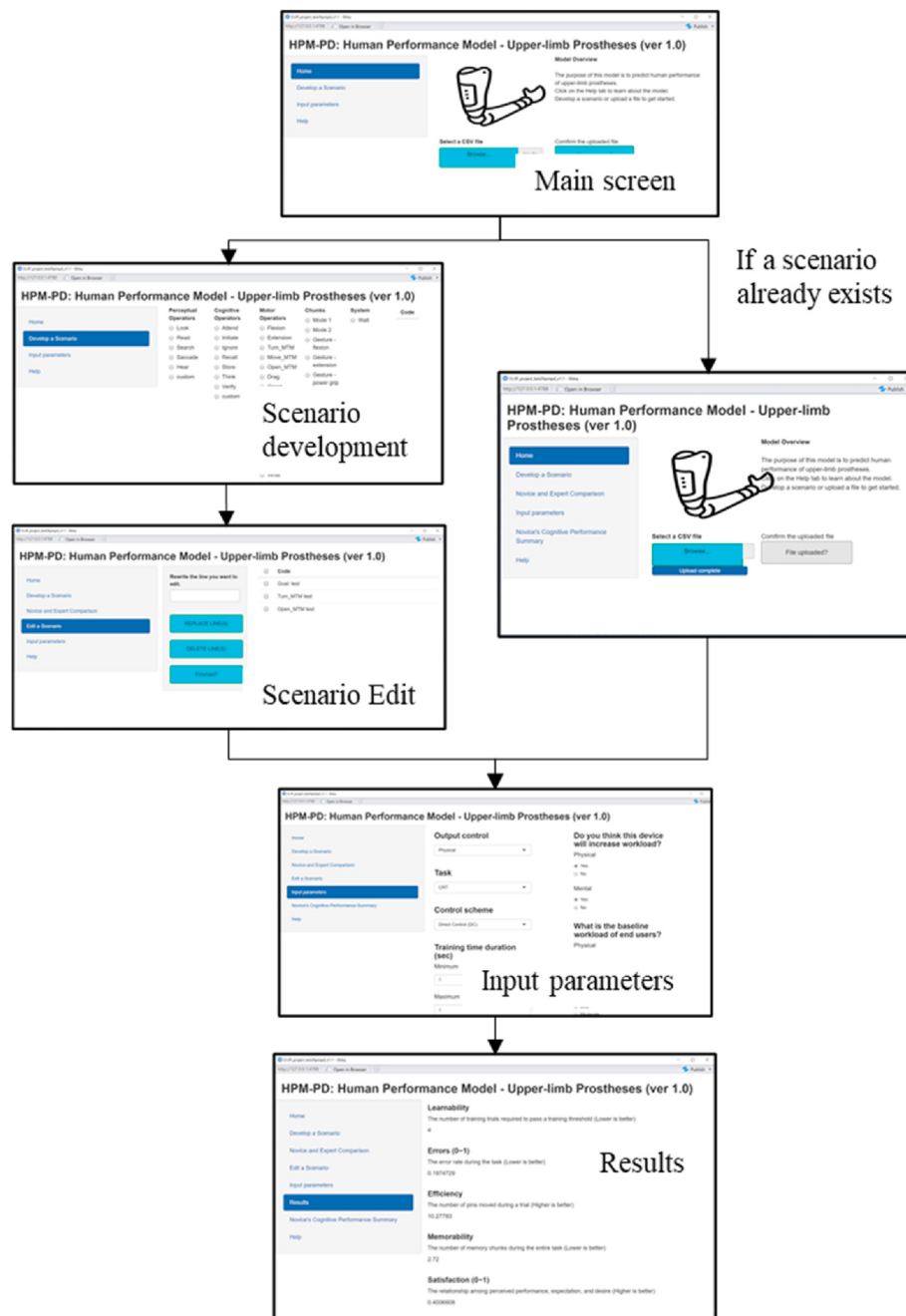


Fig. 5. Overview of HPM-UP

4. Discussion

The hypothesis test results revealed that the HPM-UP generated outcomes were similar to the outcomes derived from human subject experiment. This implies that the logic behind HPM-UP worked properly to estimate each usability dimension for upper limb prostheses under the experimental conditions of this study. However, the findings of the benchmark model were significantly different from the HPM-UP model and the outcomes from the human subject experiment. This is because the focus of the benchmark model was for modeling expert behavior without any errors. Therefore, unlike the benchmark model, HPM-UP showed closer results to the human data, especially for the TCT, as HPM-UP incorporated error rates based on the learnability dimension.

As a unique approach, in the HPM-UP, *learnability* was used as an input for calculating other usability dimensions. Including the subjective

device calibration quality and first impression in the equation was appropriate because the training criteria could not capture the individual differences. This means that although all the participants could pass the training sessions, they were not on the same level of the expertise in terms of controlling the prosthesis. Furthermore, incorporating the physical and mental workload into the learning curve slope was helpful to customize the model.

Error rate estimation feature of HPM-UP is closely related to learnability. The unique approach of HPM-UP with error rates is that it tried to model the effect of errors not only on task performance (i.e., efficiency) but also *satisfaction* because the efficiency was used as an input for the satisfaction dimension. The predicted satisfaction scores from the HPM-UP were similar to the results from human data and were significantly different from the benchmark model outcomes. The error rate calculation approach in HPM-UP was useful, as calculating the number

Home

Develop a scenario

Input parameters

Results

Help

Task

CRT

Control scheme

PR

Training time duration (sec)

Minimum

15

Maximum

25

Device calibration quality

0 0.25 0.5 0.75 1

0.67

First impression

0 0.25 0.5 0.75 1

0.5

Effort

0 0.25 0.5 0.75 1

0.31

Fig. 6. Input parameter screen.

of errors manually (by watching videos) was not feasible for the HPM-UP. It was not possible to figure out whether the specific hand gestures were errors as the videos did not provide any information on participants' intention.

We expected that the use of the CC configuration would lead to the lowest error rate followed by the PR and DC configurations. This was mainly because the CC mode allows for more intuitive and seamless transitions between different movements and has more DoF as compared to the other two control schemes (i.e., DC and PR) (Liu et al., 2024). However, this hypothesis was not supported by the data. The findings suggested that there were no significant differences in the error rate among control schemes. Although the capability of CC to drive multiple DoFs simultaneously was expected to allow participants to adopt more natural motion strategies to efficiently complete tasks, participants had a hard time controlling the device. This was because sometimes the hook was rotating even though the participants had a neutral gesture and therefore, it was difficult for them to recover from errors. The other reason might be that the ADL tasks were too simple and therefore, could not show the differences between these configurations. 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 and the task duration was short, the perceptual, cognitive, and motor demand in operating CC and PR control could be similar. There might be differences between the PR and CC configurations if the tasks become more complicated. Future studies should assess the differences in control modes using more complex ADLs.

Since error rate affects *efficiency*, the model outcomes became closer to the human subject data than the benchmark model. In a previous cognitive modeling study (Zahabi et al., 2019), it was assumed that participants were experts (which means that they do not commit errors) in a certain task after they passed the training sessions, which is one of

the main assumptions of many HPMs (Park and Zahabi, 2022b). However, this research found that participants could still make mistakes even after passing the training sessions. These errors were added to estimate the TCT and to calculate the efficiency. Therefore, the HPM-UP model outcomes were similar to the human subject data and significantly different from the benchmark model, which was developed based on CPM-GOMS and did not include error rate in the model. In sum, including error rate in the HPM-UP efficiency modeling is one of the unique merits of HPM-UP, as compared to CPM-GOMS. Another unique feature of HPM-UP is using a computational approach for quantifying *satisfaction* based on theories.

The major contribution of this research was that our model provided a method for early usability evaluation of prosthetic devices. Previous studies focused on human subject experiments and subjective evaluations for usability and mental workload assessment of amputees (Park and Zahabi, 2020, 2022a). HPM-UP can be used to improve the usability of prosthetic devices using pilot tests and early-stage prototypes especially since conducting human subject experiments with amputee participants can be time-consuming and challenging. The other contribution is that it developed a model for upper limb prosthetic devices based on top-down (theories) and bottom-up (data driven) approaches and the model was validated with a human subject experiment. The model in this research was successful as it could predict learnability accurately even for cases where participants needed more training (i.e., more than 10 trials) due to fatigue or low device calibration quality. Furthermore, the dimensions in HPM-UP are interconnected and this connectivity has not been quantified in previous methods.

4.1. Practical implications of HPM-UP

There are several practical implications of HPM-UP for clinicians,

HPM-UP: Human Performance Model – Upper-limb Prostheses

Home

Develop a scenario

Input parameters

Results

Help

Perceptual Operators	Cognitive Operators	Motor Operators	Chunks	System
☐ Look	☐ Attend	☐ Flexion	☐ Mode 1	☐ Wait
☐ Read	☐ Initiate	☒ Extension	☐ Mode 2	
☐ Search	☐ Ignore	☐ Turn	☐ Gesture 1 (flexion)	
☐ Saccade	☐ Recall	☐ Move	☐ Gesture 2 (extension)	
☐ Hear	☐ Store	☐ Open	☐ Gesture 3 (power grip)	
☐ Custom	☐ Think	☐ Drag		
	☐ Verify	☐ Grasp		
	☐ Custom	☐ Hands		
		☐ Keystroke		
		☐ Point		
		☐ Swipe		
		☐ Tap		
		☐ Touch		
		☐ Type		
		☐ Custom		

Describe the use of the operator

☐ Parallel?

☐ Add a goal?

Add new line to code

Add to current line

Remove current selections

Remove last line of code

Move to editing

Fig. 7. Scenario development screen.

device developers, or researchers in the cognitive modeling domain. The model can be used by clinicians and device developers using the GUI and with mouse-clicks. This feature can be especially useful for those without any knowledge of programming. In addition, in the 'Results' tab, a guideline table is provided, which can provide practical recommendations regarding the range of each usability dimension score. With this table, clinicians can determine whether to recommend a certain prosthesis to a patient. Finally, under the 'Help' tab of HPM-UP, several tutorial videos are provided on how to use the model.

Whenever clinicians have new amputee patients and before recommending any prosthetic device to patients, they can test or predict which device could be the best in terms of the usability for the amputee. That is, HPM-UP could reduce the work of clinicians to find, test, analyze, and recommend a prosthetic device. Once clinicians collect the input parameters (e.g., first impression) for each prosthetic device from the patients, they could run the model and see the predicted usability dimensions. Then, based on the results and the guideline table, they can recommend the best device to amputees.

For designers of prosthetic devices, HPM-UP could be a quick and practical guidance for a prototype-level usability assessment. Once they have defined the tasks and a concept for the prosthetic devices, they could predict the human performance of the device in the early stages of the design process. They can also change the input parameters based on the characteristics of the target group. Based on the results, they can make changes to the device configurations to improve human use.

Although this model was initially designed and tested with EMG-based prosthetic devices, it can be applied to assess the usability of other prosthetic device configurations (e.g., body-powered prostheses). For example, using HPM-UP for assessing the usability of the body-powered prostheses might suggest different patterns in learnability and memory load due to more intuitive and simple control mechanisms.

The model may predict a faster learning curve as the users can quickly learn to control the device through basic body movements, and it offers direct mechanical feedback, but potentially higher physical effort over long use, as compared to EMG based prostheses. In addition, the model could predict changes in usability metrics as users gain experience with the prostheses. This is because HPM-UP uses learnability as the first dimension and an input for all other following usability dimensions. Over time, users typically become more adept at operating their prosthetic devices, leading to improvements in efficiency and a reduction in error rate, which could lead to higher level of satisfaction.

4.2. Limitations

There are several aspects of this study that may limit the generalizability of findings. First, HPM-UP has some free parameters, especially in learnability and satisfaction dimensions. The reason to include them in the models was to personalize the outcome of the model to improve model performance. The initial/default values included in the current version of HPM-UP was calculated based on our pilot tests. Although this approach has been used in other HPMs such as the queuing network-model human processor (QN-MHP) (e.g., preliminary estimates of the perceptual memory access time) (Feyen, 2003), MHP (Card et al., 1986), or ACT-R (Bothell, 2020), the outcomes of the model depend on these values. To quantify qualitative dimensions such as learnability and satisfaction, some assumptions have been made. For example, *SDCQ* was calculated based on the average of the responses to questions Q3 (easiness in adjusting the device (fixing, fastening)), Q6 (easiness of using the device), and Q8 (effectiveness of using the device (the degree to which the device meets a user's needs) of the USE questionnaire. *FI* was calculated from the difference between the *SDCQ* and participant's training performance. Furthermore, HPM-UP estimates the immediate

HPM-UP: Human Performance Model – Upper-limb Prostheses

[Home](#)
[Develop a scenario](#)
[Input parameters](#)
[Results](#)
[Help](#)

Learnability
 The number of training trials required to pass training criteria (Lower is better)

Error Rate (%)
 The error rate in performing a task with a prosthetic device (Lower is better)

Efficiency (sec)
 Task completion time of one ADL cycle (e.g., moving one pin from one bar to another bar) (Lower is better)

Memory Load
 The number of memory chunks stored in working memory when performing a task with a prostheses (Lower is better)

Satisfaction (%)
 The relationship among perceived performance, expectation, and desire (Higher is better)

Fig. 8. Sample output screen of HPM-UP.

Table 3
Thresholds to interpret the findings.

Dimension	Threshold for acceptable usability	Reference
Learnability	≤ 3–5 training trials	Park et al. (2020)
Error Rate	<15%	Mohebbian et al. (2021)
Memory load	3–5 chunks of information	Cowan (2010)
Efficiency	~80% of the experts' performance	Park et al. (2022) ; White et al. (2017)
Satisfaction	Body powered device: >45% Myoelectric device: >50% Cosmetic device: >50%	Rekant et al. (2022)

learnability and satisfaction after using prostheses, which is different from the retention effect or long-term/sustained satisfaction, which is the original concept behind the ECT. Future studies should validate these assumptions with additional experiments and considering long-term satisfaction with prosthetic devices.

Table 4
Descriptive statistics (mean (sd)).

Factors (definition)	Human subject data			HPM-UP			Benchmark model		
	DC	PR	CC	DC	PR	CC	DC	PR	CC
Learnability	7.8 (3.46)	5.50 (1.50)	9.80 (4.40)	8.20 (6.54)	7.00 (3.87)	10.00 (5.00)	N/A		
Error rate (%)	N/A			36.43 (3.36)	25.8 (3.55)	28.14 (8.76)	N/A		
Memory Load	N/A			4.79 (0.04)	0	0	4.80 (0.00)	0	0
Efficiency CRT	17.91 (6.76)	12.90 (5.77)	13.40 (2.30)	19.03 (2.48)	11.19 (3.91)	14.19 (5.62)	13.3 (0.00)	6.9 (0.00)	6.9 (0.00)
SHAP	11.03 (2.31)	11.26 (4.95)	18.91 (7.10)	11.96 (0.70)	12.22 (1.42)	13.18 (1.71)	8.2 (0.00)	8.6 (0.00)	8.6 (0.00)
Satisfaction (%)	75.09 (0.12)	64.51 (0.23)	65.28 (0.18)	61.16 (0.06)	61.18 (0.11)	56.20 (0.16)	N/A		

Table 5
Summary of hypothesis test results.

Hypothesis ID	Hypothesis	Test Result	Test statistics, p-value, effect size
H1 (Learnability)	The results of HPM-UP learnability dimension would be similar to the human-subject data.	Supported	$W = 436.5, p = 0.85, r = 0.03$
H2 (Error rate)	Use of the CC configuration would lead to the lowest error rate followed by the PR and DC configurations.	Refuted (DC = PR=CC)	$H(2) = 1.57, p = 0.47, \eta^2 = .06$
H3 (Memory load)	The results of HPM-UP memory load dimension would be similar to the benchmark model	Supported	$W = 65, p = 0.23, r = 0.28$
H4 (Efficiency)	(H4-1) The results of HPM-UP efficiency dimension would be similar to the human subject data	Supported	$Z = 0.26, p = 0.79, r = .03$
	(H4-2) There would be a significant difference between the HPM-UP efficiency dimension results and the benchmark model results	Supported	$Z = -4.54, p < 0.001, r = 0.64$
H5 (Satisfaction)	The results of HPM-UP satisfaction dimension would be similar to the human-subject data	Supported	$W = 41.3, p = 0.59, r = 0.07$

usability assessments. Future studies should validate HPM-UP with participants from different age groups.

The decision to work with healthy individuals 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. Furthermore, although HPM-UP provides estimates of device usability, it cannot guarantee the fitness or feeling of embodiment of a prosthesis to amputees. Lastly, although HPM-UP supports modeling under GUI environment, analysts need to have basic knowledge of human performance modeling.

5. Conclusion

Previous studies for measuring usability of upper-limb prostheses relied on human subject experiments and subjective evaluations. While these methods could provide useful outcomes, early estimation of usability is critical to reduce future device rejection due to usability issues. Therefore, this research advanced the fundamental knowledge of estimating usability of EMG-based upper-limb prostheses. This study not only quantified each usability dimension (learnability, errors, memory load, efficiency, and satisfaction) but also connected them in a computational way. The outcome of this study, HPM-UP, provides the capability to predict human performance of prostheses at the early stage of the design process.

The model is released to Github (<https://github.com/hsilab/hpmup>) and can be run with Rstudio. As the model was developed in R Shiny package format with a GUI, it does not require hard coding and other researchers, designers, or clinicians can easily adopt it. As the source code is released, researchers can modify or upgrade the software for their own purposes. Clinicians can test and analyze the human performance of several commercial prostheses to find and recommend a best device(s) for the patient. Manufacturers can also run the software to evaluate usability of devices under different types of prosthetic configurations and scenarios.

Disclosure of interest

None.

Declaration of interest statement

Research article entitled “A novel approach for predicting usability of upper limb prostheses” has no conflict of interest to declare.

CRediT authorship contribution statement

Junho Park: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maryam Zahabi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project

administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **He Huang:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Mark Benden:** Writing – review & editing, Writing – original draft, Supervision, Investigation.

Acknowledgments

This project was funded by the National Science Foundation (No. IIS-1856676/1856441). The opinions expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Bhattacharjee, A., 2001. Understanding information systems continuance: an expectation-confirmation model. *MIS Q.* 351–370.
- Biddiss, E.A., Chau, T.T., 2007. Upper limb prosthesis use and abandonment: a survey of the last 25 years. *Prosthet. Orthot. Int.* 31 (3), 236–257. <https://journals.sagepub.com/doi/pdf/10.1080/03093640600994581>.
- Bothell, D., 2017. ACT-R 7 Reference Manual. Available at: act-r.psy.cmu.edu/wordpress/wpcontent/themes/ACT-R/actr7/reference-manual.pdf.
- Bothell, D., 2020. ACT-R 7.21+ Reference Manual.
- Bravini, E., Franchignoni, F., Ferriero, G., Giordano, A., Bakhsh, H., Sartorio, F., Vercelli, S., 2014. Validation of the Italian version of the client satisfaction with device module of the Orthotics and prosthetics users' survey. *Disability and health journal* 7 (4), 442–447.
- Brooke, J., 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189 (194), 4–7.
- Camm, J.D., 1985. A note on learning curve parameters. *Decis. Sci. J.* 16 (3), 325–327.
- Card, S.K., Moran, T.P., Newell, A., 1986. The model human processor- an engineering model of human performance. *Handbook of perception and human performance* 2, 1–35.
- Cowan, N., 2010. The magical mystery four: how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* 19 (1), 51–57.
- Demers, L., Weiss-Lambrou, R., Ska, B., 2002. The Quebec user evaluation of satisfaction with assistive technology (QUEST 2.0): an overview and recent progress. *Technol. Disabil.* 14 (3), 101–105.
- Desrosiers, J., Bravo, G., Hébert, R., Dutil, É., Mercier, L., 1994. Validation of the Box and Block Test as a measure of dexterity of elderly people: reliability, validity, and norms studies. *Arch. Phys. Med. Rehabil.* 75 (7), 751–755.
- Dix, A., Finlay, J., Abowd, G.D., Beale, R., 2000. *Human-computer Interaction*. Harlow ua.
- Dunn, O.J., 1964. Multiple comparisons using rank sums. *Technometrics* 6 (3), 241–252.
- Engdahl, S.M., Christie, B.P., Kelly, B., Davis, A., Chestek, C.A., Gates, D.H., 2015. Surveying the interest of individuals with upper limb loss in novel prosthetic control techniques. *J. NeuroEng. Rehabil.* 12 (1), 53.
- Estes, S., 2015. The workload curve: subjective mental workload. *Hum. Factors* 57 (7), 1174–1187.
- Estes, S., 2017. Cogulator. *The MITRE Corporation*.
- Estes, S., 2021. Cogulator - A Cognitive Modeling Calculator. <https://github.com/Cogulator/Cogulator>.
- Feyen, R.G., 2003. Modeling Human Performance Using the Queuing Network-Model Human Processor (QN-MHP).
- Fontana, J.M., 2010. Classification of EMG Signals to Control a Prosthetic Hand Using Time-Frequency Representations and Support Vector Machines. Louisiana Tech University.
- Gao, M., Kortum, P., Oswald, F., 2018. Psychometric evaluation of the use (usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.
- Hargrove, L., Losier, Y., Lock, B., Englehart, K., Hudgins, B., 2007. A real-time pattern recognition based myoelectric control usability study implemented in a virtual

- environment. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.
- Hart, S.G., 2006. October). NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9). SAGE Publications, Los Angeles, CA, pp. 904–908.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load Index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. Elsevier.
- ISO, 2019. Ergonomics of Human-System Interaction-Part 11: Usability: Definitions and Concepts (ISO 9241-11: 2018) Irish Standard ISO. *Recuperado de*. infostore.saiglobal.com/preview/is/en/2018/is_eniso9241-11-2018.pdf.
- John, B.E., Gray, W.D., 1994. GOMS analysis for parallel activities. In: Conference Companion on Human Factors in Computing Systems.
- John, B.E., Kieras, D.E., 1994. In: The GOMS Family of Analysis Techniques: Tools for Design and Evaluation.
- Joyce, A., 2019. How to Measure Learnability of a User Interface. <https://www.nngroup.com/articles/measure-learnability/>.
- Kieras, D.E., Santoro, T.P., 2004. Computational GOMS modeling of a complex team task: lessons learned. *Conference on human Factors in computing systems - proceedings*. 2004 Conference on Human Factors in Computing Systems - Proceedings, CHI 2004. April 24, 2004 - April 29, 2004, Vienna, Austria.
- Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* 47 (260), 583–621.
- Leiden, K., Best, B., 2005. A Cross-Model Comparison of Human Performance Modeling Tools Applied to Aviation Safety, vol. 80301. Micro Analysis & Design, Inc., Boulder, CO, p. 2005.
- Liu, Y., Berman, J., Dodson, A., Park, J., Zahabi, M., Huang, H., Ruiz, J., Kaber, D.B., 2024. Human-centered evaluation of EMG-based upper-limb prosthetic control modes. *IEEE Trans. Hum.-Mach. Syst.* 54 (3), 271–281.
- Lock, B., Englehart, K., Hudgins, B., 2005. Real-time Myoelectric Control in a Virtual Environment to Relate Usability vs. Accuracy.
- Lowry, P.B., Spaulding, T., Wells, T., Moody, G., Moffit, K., Madariaga, S., 2006. A theoretical model and empirical results linking website interactivity and usability satisfaction. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06).
- Lund, A.M., 2001. Measuring Usability with the Use Questionnaire 12, vol. 8. Usability interface, pp. 3–6 (2).
- Maresova, P., Krejcar, O., Maskuriy, R., Bakar, N.A.A., Selamat, A., Truhlarova, Z., Horak, J., Joukl, M., Vitkova, L., 2023. Challenges and opportunity in mobility among older adults—key determinant identification. *BMC Geriatr.* 23 (1), 447.
- Maynard, H.B., Stegemerten, G.J., Schwab, J.L., 1948. Methods-time Measurement.
- McLellan, S., Muddimer, A., Peres, S.C., 2012. The effect of experience on system usability scale ratings. *Journal of usability studies* 7 (2), 56–67.
- Mislik, G.K., Nussbaum, D.A., 2015. Cost Estimation: Methods and Tools. John Wiley & Sons.
- Mohebbian, M.R., Nosouhi, M., Fazilati, F., Esfahani, Z.N., Amiri, G., Malekifar, N., Yusefi, F., Rastegari, M., Marateb, H.R., 2021. A Comprehensive Review of Myoelectric Prosthesis Control arXiv preprint arXiv:2112.13192.
- Montagnani, F., Controzzi, M., Cipriani, C., 2015. Is it finger or wrist dexterity that is missing in current hand prostheses? *IEEE Trans. Neural Syst. Rehabil. Eng.* 23 (4), 600–609. <https://doi.org/10.1109/tnsre.2015.2398112>.
- Music, C.A., 2022. ELECTROMYOGRAPHY-BASED ASSISTIVE VIRTUAL REALITY HUMAN-MACHINE INTERFACE. UNIVERSITY OF FLORIDA.
- Niamba, K., Schieber, F., McCray, M., 2021. Myoelectric control: an alternative to mirror therapy. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 65, No. 1). SAGE Publications, Los Angeles, CA, pp. 948–950.
- Nielsen, J., 2012. Usability 101: Introduction to Usability.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9 (1), 97–113.
- Oliver, R.L., 1977. Effect of expectation and disconfirmation on postexposure product evaluations: an alternative interpretation. *J. Appl. Psychol.* 62 (4), 480.
- Oliver, R.L., 1980. A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Market. Res.* 17 (4), 460–469.
- Parajuli, N., Sreenivasan, N., Bifulco, P., Cesarelli, M., Savino, S., Niola, V., Esposito, D., Hamilton, T.J., Naik, G.R., Gunawardana, U., 2019. Real-time EMG based pattern recognition control for hand prostheses: a review on existing methods, challenges and future implementation. *Sensors* 19 (20), 4596.
- Park, J., 2023. *Human Performance Modeling Of Upper Limb Prosthetic Devices* [Doctoral Dissertation. Texas A&M University].
- Park, J., Berman, J., Dodson, A., Liu, Y., Armstrong, M., Huang, H., Kaber, D., Ruiz, J., Zahabi, M., 2023. Assessing workload in using electromyography (EMG)-based prostheses. *Ergonomics* 1–24.
- Park, J., Berman, J., Dodson, A., Liu, Y., Matthew, A., Huang, H., Kaber, D., Ruiz, J., Zahabi, M., 2022. Cognitive workload classification of upper-limb prosthetic devices. In: 2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS).
- Park, J., Zahabi, M., 2020. Comparison of cognitive workload assessment techniques in EMG-based prosthetic device studies. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC).
- Park, J., Zahabi, M., 2022a. Cognitive workload assessment of prosthetic devices: a review of literature and meta-analysis. In: *Ieee Transactions on Human-Machine Systems*.
- Park, J., Zahabi, M., 2022b. A review of human performance models for prediction of driver behavior and interactions with in-vehicle technology. *Hum. Factors*, 00187208221132740.
- Park, J., Zahabi, M., Kaber, D., Ruiz, J., Huang, H., 2020. Evaluation of activities of daily living tesbeds for assessing prosthetic device usability. In: 2020 IEEE International Conference on Human-Machine Systems (ICHMS).
- Peerdeman, B., Boere, D., Witteveen, H., Hermens, H., Stramigioli, S., Rietman, H., Veltink, P., Misra, S., 2011. Myoelectric forearm prostheses: state of the art from a user-centered perspective. *J. Rehabil. Res. Dev.* 48 (6), 719–738.
- Pinheiro, M.B., Oliveira, J.S., Baldwin, J.N., Hassett, L., Costa, N., Gilchrist, H., Wang, B., Kwok, W., Albuquerque, B.S., Pivotto, L.R., 2022. Impact of physical activity programs and services for older adults: a rapid review. *Int. J. Behav. Nutr. Phys. Activ.* 19 (1), 87.
- Rekant, J., Fisher, L.E., Boninger, M.L., Gaunt, R.A., Collinger, J.L., 2022. Amputee, clinician, and regulator perspectives on current and prospective upper extremity prosthetic technologies. *Assist. Technol.* 1–13.
- Resnik, L., Huang, H.H., Winslow, A., Crouch, D.L., Zhang, F., Wolk, N., 2018. Evaluation of EMG pattern recognition for upper limb prosthesis control: a case study in comparison with direct myoelectric control. *J. NeuroEng. Rehabil.* 15 (1), 23.
- Rosenthal, R., 1986, 1984, 148 pp. Educational Researcher. Meta-Analytic Procedures for Social Science Research, vol. 15. Sage Publications, Beverly Hills, pp. 18–20 (8).
- Sanders, J.E., Fatone, S., 2011. Residual limb volume change: systematic review of measurement and management. *J. Rehabil. Res. Dev.* 48 (8), 949.
- Scheme, E., Englehart, K., 2011. Electromyogram pattern recognition for control of powered upper-limb prostheses: state of the art and challenges for clinical use. *J. Rehabil. Res. Dev.* 48 (6).
- Spreng, R.A., MacKenzie, S.B., Olshavsky, R.W., 1996. A reexamination of the determinants of consumer satisfaction. *J. Market.* 60 (3), 15–32.
- Tiffin, J., Asher, E.J., 1948. The Purdue Pegboard: norms and studies of reliability and validity. *J. Appl. Psychol.* 32 (3), 234.
- Toledo, D.R., Barela, J.A., 2014. Age-related differences in postural control: effects of the complexity of visual manipulation and sensorimotor contribution to postural performance. *Exp. Brain Res.* 232, 493–502.
- Tomczak, M., Tomczak, E., 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in sport sciences* 21 (1).
- VandenBos, G.R., 2007. APA Dictionary of Psychology. American Psychological Association.
- White, M.M., Zhang, W., Winslow, A.T., Zahabi, M., Zhang, F., Huang, H., Kaber, D.B., 2017. Usability comparison of conventional direct control versus pattern recognition control of transradial prostheses. *Ieee Transactions on Human-Machine Systems* 47 (6), 1146–1157.
- Wilcoxon, F., 1992. Individual comparisons by ranking methods. In: *Breakthroughs in Statistics*. Springer, pp. 196–202.
- Wu, C., 2018. The five key questions of human performance modeling. *Int. J. Ind. Ergon.* 63, 3–6.
- Yeh, Y.-Y., Wickens, C.D., 1988. Dissociation of performance and subjective measures of workload. *Hum. Factors* 30 (1), 111–120.
- Zahabi, M., White, M.M., Zhang, W., Winslow, A.T., Zhang, F., Huang, H., Kaber, D.B., 2019. Application of cognitive task performance modeling for assessing usability of transradial prostheses. *Ieee Transactions on Human-Machine Systems* 49 (4), 381–387. <https://doi.org/10.1109/THMS.2019.2903188>.
- Zhang, W., Ma, W., Brandao, M., Kaber, D.B., Bloomfield, P., Swangnetr, M., 2016. Biometric validation of a virtual reality-based psychomotor test for motor skill training. *Assist. Technol.* 28 (4), 233–241.