



## Advanced modeling method for quantifying cumulative subjective fatigue in mid-air interaction

Ana Villanueva<sup>a,1,\*</sup>, Sujin Jang<sup>d,1</sup>, Wolfgang Stuerzlinger<sup>b</sup>, Satyajit Ambike<sup>c</sup>, Karthik Ramani<sup>a</sup>

<sup>a</sup> School of Mechanical Engineering, Purdue University, West Lafayette, IN, USA

<sup>b</sup> School of Interactive Arts + Technology, Simon Fraser University, Vancouver, CA, Canada

<sup>c</sup> Department of Health and Kinesiology, Purdue University, West Lafayette, IN, USA

<sup>d</sup> Samsung Advanced Institute of Technology, Suwon-si, Gyeonggi-do, South Korea

### ARTICLE INFO

#### Keywords:

Mid-air interaction  
Cumulative fatigue model  
Maximum arm strength  
Brain effort

### ABSTRACT

Interaction in mid-air can be fatiguing. A model-based method to quantify cumulative subjective fatigue for such interaction was recently introduced in HCI research. This model separates muscle units into three states: active ( $M_A$ ) fatigued ( $M_F$ ) or rested ( $M_R$ ) and defines transition rules between states. This method demonstrated promising accuracy in predicting subjective fatigue accumulated in mid-air pointing tasks. In this paper, we introduce an improved model that additionally captures the variations of the maximum arm strength based on arm postures and adds linearly-varying model parameters based on current muscle strength. To validate the applicability and capabilities of the new model, we tested its performance in various mid-air interaction conditions, including mid-air pointing/docking tasks, with shorter and longer rest and task periods, and a long-term evaluation with individual participants. We present results from multiple cross-validations and comparisons against the previous model and identify that our new model predicts fatigue more accurately. Our modeling approach showed a 42.5% reduction in fatigue estimation error when the longitudinal experiment data is used for an individual participant's fatigue. Finally, we discuss the applicability and capabilities of our new approach.

### 1. Introduction

Fatigue affects the force-production capacity of muscles, as well as inter-joint and inter-muscular coordination. Arm fatigue has been cited as risk factor for work-related injury (Kinali et al., 2016). While arm fatigue is associated with repetitive injury, the underlying mechanisms remain poorly understood. However, the advancement of mathematical modeling may help in more accurately quantifying muscle fatigue development (Loofit et al., 2018). Recently, arm fatigue has become an important factor in the design of human-computer interfaces (HCI) (Bachynskyi et al., 2015; Hincapié-Ramos et al., 2014). As AR/VR technologies such as Oculus (Anon, 2020d,b,c), Hololens 2 (Anon, 2020a), etc, become ubiquitous, HCI will require a comprehensive, robust, and practical method for predicting arm fatigue. A method that considers both physiological and psychological factors that contribute to fatigue (Enoka and Stuart, 1992) and assesses fatigue without interference with the ongoing activity is required. Assessment of subjective fatigue is important, as excess subjective fatigue in industrial settings may be a biomarker of increased injury risk, and it negatively impacts user experience in HCI (Hassenzahl and Tractinsky, 2006). An interference-free measurement method could be used to

assess fatigue both in the workplace and in exergames, to avoid long-term deficits such as WMSDs, evaluate novel interaction techniques, optimize user input positions and orientations, adjust the range of motion, and avoid uncomfortable postures for the user.

Objective fatigue evaluations that measure physiological quantities such as muscle activation (Cifrek et al., 2009), heart rate (Segerstrom and Nes, 2007), blood pressure (Sjøgaard et al., 1988), and blood oxygen level (Amann et al., 2006) require specialized equipment and interfere with the ongoing activity. Similarly, subjective fatigue measurements like the Likert scale (Carifio and Perla, 2007), the NASA-TLX (Hart and Staveland, 1988), the VAS (visual analog signal) (Bijur et al., 2001), ratings of perceived exertion (RPE) (Borg, 1982), and the Borg CR10 scales (Hincapié-Ramos et al., 2014; Jang et al., 2017) require repeated verbalization of fatigue levels and thus, interfere with the ongoing activity.

Users of mid-air interface typically perceive accumulation of fatigue (i.e., feel the arm getting heavier) over time, and this perception may negatively impact the user experience. Further, subjective fatigue and its perceived intensity are influenced by the demand of the task, the user's physiological state, and the user's history of

\* Correspondence to: School of Mechanical Engineering, 585 Purdue Mall, Purdue University, West Lafayette, IN 47907, USA.  
E-mail address: [villana@purdue.edu](mailto:villana@purdue.edu) (A. Villanueva).

<sup>1</sup> Both authors contributed equally to this work.

movement choices (Jang et al., 2017). Although there are model-based approaches that quantify cumulative fatigue based on physical (e.g., forces, torques) or physiological (e.g., muscle activation, heart rate) measurements (Xia and Law, 2008), there is relatively little work done model-based estimation of subjective fatigue. Further, perceived fatigue may vary during static or dynamic tasks. Law et al. (2010). To address these gaps, we recently used the three-compartment muscle (TCM) model (Xia and Law, 2008) and developed a method to predict cumulative subjective fatigue for mid-air pointing tasks (Jang et al., 2017). Once the estimation models are created through a calibration process, our method estimates subjective fatigue using only remote measurement of the movement kinematics, thereby eliminating any interference with the ongoing activity. The method does not require expensive equipment, and furthermore, allows the quantification of cumulative fatigue by accounting for both task and rest periods, closely mimicking natural behaviors.

This paper presents an improved version of our earlier model. The enhancements include: (1) an improved estimation of maximal shoulder torque; (2) incorporation of *brain effort* (BE) as a proxy for subjective fatigue; (3) experimental validation of the improved model on (a) more complex mid-air manual tasks, (b) tasks with varying durations of rest/task periods, and personalized models for long-term tasks performed over several days.

This model separates muscle units into three states: active ( $M_A$ ), fatigued ( $M_F$ ), or rested ( $M_R$ ), and defines transition rules between states.

The TCM model provides a method that takes both physical and subjective measurements (i.e., perception of muscle fatigue). It involves non-linear differential equations to describe the flow between compartments as the product of the constant parameters times the volume of the compartment. However, the actual flow rate is not constant, but inherently changing as a function of the compartments. Thus, the flow rates could be different based on the current muscle capacity and task load (Fuglevand et al., 1993). Further, Xia and Frey Law (Xia and Law, 2008) included brain effort (BE) into the TCM model as the representation of the central drive required to perform a task. The BE term was adopted to quantify fatigue during dynamic load tasks. In prior work, BE was used for physiological fatigue evaluation (e.g., reduction of muscle strength) but not for subjective fatigue quantification. To capture the subjective and physiological changes in fatiguing and resting rates, we expand the TCM model to include the BE term into our approach. Subsequently, we expand the practicality of this model-based approach by an improvement in fatigue estimation performance. Moreover, we validate the applicability of our approach in various interaction conditions and more complex tasks (i.e., situations beyond a simple pointing task).

Maximum shoulder torque, which varies with joint angle and angular velocity, is an important input to quantify fatigue status. The TCM model assumes that maximum joint torque depends on individual physical capabilities. For instance, the posture of the arm is also known to affect the maximum shoulder torque (Coury et al., 1998). The maximum power capacity of arm muscles when we stretch the arm away from the body could be different from when we are at a natural standing pose. Thus, we adopt Chaffin's model of maximum shoulder torque, which is based on elbow and shoulder angles (Chaffin et al., 1999), to get posture-based maximum shoulder torque estimates.

Jang et al. (2017) validated the TCM model through a simple mid-air pointing task under incremental subjective fatigue accumulation. Their evaluation also did not consider task and rest periods of varying length. For more general application scenarios and to validate the applicability of the model-based approach in estimating subjective fatigue, varying durations of task periods and rest should be taken into account. Moreover, realistic mid-air interaction commonly includes not only pointing tasks, but also more complex interaction, such as 3D docking tasks, which combine rotation and translation (Vuibert et al., 2015). In the work presented here, we design a series of experiments to

further validate the feasibility of the model-based approach to fatigue prediction for various interaction conditions. Also, we compare the performance of our proposed fatigue model with existing methods.

Our contributions include: (1) a specific maximum strength representation compatible with the TCM fatigue modeling method that includes posture-based maximum shoulder torque estimation; (2) connecting subjective fatigue and muscle fatigue without contact-based measurement; (3) a reliable cumulative fatigue model based on brain effort (BE); (4) an experimental validation of the estimation performance of the modified model during a complex mid-air task, varying durations of rest/task periods, and individual long-term task conditions; (5) a 42.5% decrease in fatigue estimation error for individualized fatigue modeling.

## 2. Related work

Our work broadly links to fatigue estimation methods, maximum arm strength estimation, and arm fatigue in HCI.

### 2.1. Objective and subjective fatigue evaluation

Objective fatigue evaluation methods involve direct measurements of various physiological quantities, such as muscle activation (Cifrek et al., 2009), heart rate (Segerstrom and Nes, 2007), blood pressure (Sjøgaard et al., 1988), and blood oxygen level (Amann et al., 2006). However, these approaches require specialized, fully calibrated equipment and are generally too invasive to be used as part of common user interfaces. Subjective fatigue measurements include the Likert scale (Carifio and Perla, 2007), the NASA-TLX (Hart and Staveland, 1988), the VAS (visual analog signal) (Bijur et al., 2001), and ratings of perceived exertion (RPE) (Borg, 1982). In general, the Borg CR10 scales have been preferred in HCI research, due to their strong correlation with arm fatigue (Hincapié-Ramos et al., 2014; Jang et al., 2017), and physiological measurements (e.g., electromyography) under light load (or bare hand) interaction conditions (Öberg et al., 1994). Such subjective measurements provide critical assessments of the user experience during interaction. However, these methods require repeated verbalization of fatigue levels during interaction. Thus, direct measurements of subjective fatigue could interfere with interaction tasks.

In contrast, our model-based approach does not require any expensive set-up nor repeated measurements after we generate the estimation models, which account for physical and subjective influences on fatigue.

### 2.2. Modeling method to muscle fatigue quantification

The three-compartment muscle (TCM) model assumes muscle units can be in either one of active ( $M_A$ ), fatigue ( $M_F$ ), and rest ( $M_R$ ) states (Xia and Law, 2008). Recently, Jang et al. (2017) validated the applicability of the TCM model in estimating cumulative subjective fatigue based on movement kinematics.

The TCM model assumes constant parameters in the three differential equations, but the modeling of muscle responses requires nonlinear approaches (Fuglevand et al., 1993). Sonne and Potvin (2016) proposed a modification of the model based on a relationship between fatigue/resting rate and target task load (Bigland-Ritchie et al., 1986). However, their work targeted the quantification of fatigue level based on a reduction in direct capacity measurements, such as grasping force and torque. Other works (Frey-Law et al., 2021; Looft and Frey-Law, 2020; Rashedi and Nussbaum, 2015; Gede and Hubbard, 2014) have also proposed modified models that evaluate what is lacking in the original TCM model (e.g., how the number and placement of fatigue data influences parameter identification), but they do not approach modeling subjective fatigue with a non-contact approach. Thus, in line with Sonne et al.'s work, we propose a modification of the TCM model by assuming a linear relationship between fatigue/resting parameters as a function of BE and the target load. In our experiments, we investigate the validity and applicability of the assumption in estimating cumulative subjective fatigue.

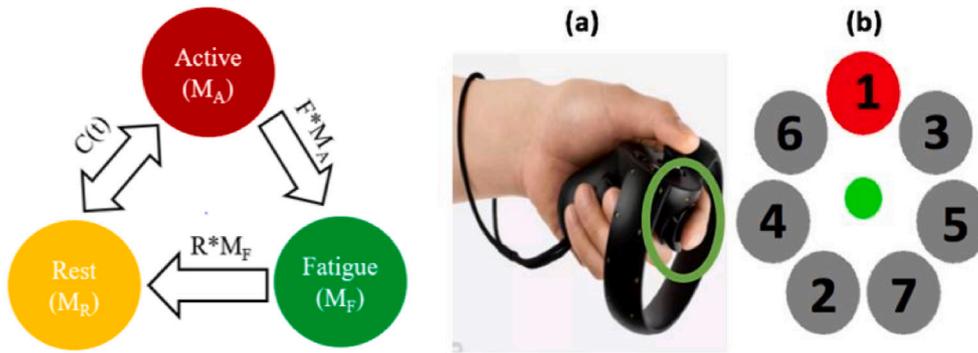


Fig. 1. Left: Overview of the three-compartment model (TCM), which represents the percentages of motor units that are at rest ( $M_R$ ), active ( $M_A$ ), and fatigue ( $M_F$ ) states.  $F$  and  $R$  represent the fatigue and recovery coefficients, and  $C(t)$  defines the muscle activation–deactivation drive. Right: Controller for pointing task and targets. (a): Oculus Touch device used to control the pointer (green dot in (b)) and button (green circle in (a)) to indicate “selection”, (b): target circles (current target in red, inactive in gray) having equal distance to each other. Target width is 10 cm and distance is 30 cm. The number indicates the selection order. Numbers were not visible to participants.

### 2.3. Maximum arm strength estimation

Enoka and Stuart (1992) defined fatigue as a reduction of muscle strength relative to the maximal value. Thus, any reliable arm fatigue evaluation requires an individual’s maximum arm strength defined as the *maximum voluntary contraction* (MVC). In biomechanics research, specialized test rigs have been commonly used for measuring MVC in torque or force units (Hayes et al., 2002; Roy et al., 2011). However, these methods are expensive and impractical to be used in many user interfaces. In recent years, Jang et al. (2017) proposed a simple but effective method to estimate the maximum shoulder strength through an isometric load task. Although this method showed a strong correlation with traditional contact-based measurements, it ignores the variation of arm strength based on the arm posture. Yet, it is well known that the arm strength varies based on the person’s posture (Coury et al., 1998). Based on previous work, we use Chaffin’s strength model (Chaffin et al., 1999) for biomechanically reliable fatigue estimation, which estimates the maximal shoulder torque (MVC) based on elbow and shoulder joint angles as well as biological sex.

### 2.4. Arm fatigue evaluation in HCI

Recently, researchers investigated arm fatigue and its quantification in HCI. Bachynskyi et al. (2014) introduced a biomechanical simulation method to estimate muscle activations. Consumed Endurance (CE) (Hincapié-Ramos et al., 2014) is a fatigue metric that showed a strong correlation with subjective fatigue measures, more specifically, Borg ratings (Borg, 1982). These methods were used to study the impact of arm fatigue in mid-air (Hincapié-Ramos et al., 2014) and various touch interface designs (Bachynskyi et al., 2015). However, these approaches cannot quantify the effect of rest on the accumulation of fatigue. Moreover, prior methods ignored individual differences in arm strength.

Jang et al. (2017) proposed a model-based method, based on the TCM model, to capture the effect of rest on cumulative subjective fatigue based on individual arm strength. This method was validated in a simple pointing task under incremental fatiguing condition. Also, this method assumes constant fatiguing and resting rates. However, the rate of muscle fatiguing and relaxation could decrease or increase based on task types and current exertion level (Enoka and Stuart, 1992). To further improve the physiological fidelity of the model-based method, we modified the fatigue model parameters by defining a linear relationship between current muscular capacity and fatigue parameters. We also designed experiments to test the applicability and capabilities of our new approach.

### 3. Quantifying cumulative fatigue

In this section, we describe our new, modified cumulative fatigue model and maximum arm strength estimation method.

#### 3.1. Modified three-compartment fatigue model

The three-compartment muscle (TCM) model (Fig. 1) assumes motor units can be in either one of active ( $M_A$ ), fatigue ( $M_F$ ), and rest ( $M_R$ ) states (see Xia and Law (2008), Frey-Law et al. (2012), Jang et al. (2017) for more details). Each compartment of motor states is expressed as a percentage of maximum voluntary contraction (% of MVC). The sum of each compartment is 100%, as our muscle motor unit quantity does not suddenly change during tasks. Since motor-unit recruitment is binary, this means that a motor unit is either contracted or it is not. For a MVC task, all motor units are contracted and for a sub-maximal task, fewer motor units are contracted. The transition among motor units is defined as:

$$\begin{aligned} \frac{dM_R}{dt} &= -C(t) + R * M_F, \\ \frac{dM_A}{dt} &= C(t) - F * M_A, \\ \frac{dM_F}{dt} &= F * M_A - R * M_F, \end{aligned}$$

where  $F$  and  $R$  defines the rate of motor state transitions activation–fatigue and fatigue–rest.  $C(t)$  is motor unit activation function defined as:

$$C(t) = \begin{cases} L_D(TL - M_A) & \text{if } M_A < TL, M_R > TL - M_A \\ L_D M_R & \text{if } M_A < TL, M_R \leq TL - M_A \\ L_R(TL - M_A) & \text{if } M_A \geq TL \end{cases}$$

$TL$  is the target load defined as a torque ratio  $[T_{current}/T_{max}] * 100\%$ ,  $L_D$  is the force development rate, and  $L_R$  is the relaxation factor. The last two parameters are set to 10 based on the sensitivity analysis by Frey-Law et al. (2012). Residual capacity ( $RC$ ) indicates the current muscle strength, which is defined as:

$$RC = M_A + M_R = 100\% - M_F$$

Based on the residual capacity, Xia and Law (2008) proposed a *Brain effort* ( $BE$ ) term that defines the required central “drive” to perform a task:

$$BE = \begin{cases} TL/RC * 100\% & \text{if } TL \leq RC \\ 100\% & \text{if } TL > RC \end{cases}$$

where  $BE$  can be related with participant’s subjective fatigue. The  $BE$  ranges between 0 and 0.638. Sonne and Potvin (2016) proposed a modified TCM model (graded motor unit, GMU) to evaluate fatigue

based on physical measurements, e.g., reduction in maximum grasping force, by assuming a linear relationship between the fatiguing rate  $F$  and  $BE$ :  $F = F_c * BE$  where  $F_c$  is a constant value. They proposed a modified resting rate  $R$  as:  $R = R_c * BE * M_F * (\alpha - BE)/\alpha$ , where  $\alpha$  is a threshold of 63.8%. In our preliminary evaluation of this model in estimating cumulative subjective fatigue, GMU showed lower estimation accuracy than the TCM model (RMSE-GMU = 2.03, RMSE-TCM = 1.80). The threshold value is derived from physiological muscle measurements (i.e., electromyography) under an isometric load condition (i.e., fixed load). To the best of our knowledge, we are not aware of work supporting such a threshold value in subjective fatigue evaluation. Also, our experimental condition does not involve constant but dynamically changing arm movements. Based on this observation, we investigated a modified model (LIN) based on the linear definition of the fatiguing/resting rates ( $F$ ,  $R$ ) as a function of  $BE$  and target load:

$$F = F_s * BE + F_b$$

$$R = R_s * BE + R_b$$

where  $F_s$  and  $R_s$  are constants defining the effect of  $BE$  on the rates and  $F_b$  and  $R_b$  are constants defining the rates when  $BE$  is zero (i.e., resting condition,  $TL = 0\%$ ).

### 3.1.1. Model fitting

We used the optimization toolbox in MATLAB to optimize the TCM model and to find the model parameters ( $F$  and  $R$ ) in mid-air interaction. To find the rate parameters  $F_s$ ,  $F_b$ ,  $R_s$ , and  $R_b$  for an optimal model performance in estimating cumulative subjective fatigue, we formalized an error function as:

$$\text{minimize}_{F_s, F_b, R_s, R_b} \sqrt{\frac{1}{n} \sum_{i=1}^n [\phi(M_F(i)) - B(i)]^2}$$

participant to  $F_s \in [F_{s_{lb}}, F_{s_{ub}}]$ ,  $F_b \in [F_{b_{lb}}, F_{b_{ub}}]$ ,  $R_s \in [R_{s_{lb}}, R_{s_{ub}}]$ , and  $R_b \in [R_{b_{lb}}, R_{b_{ub}}]$ . In the above equation,  $n$  is the number of fitting data points,  $M_F(i)$  is the fatigue level estimation, and  $B(i)$  is the Borg scale rating. Based on prior work (Jang et al., 2017; Morishita et al., 2014), the scaling function  $\phi(x)$  is defined as:  $\phi(x) = 0.0875 * x$ . Due to the discontinuity in the model functions, finding optimal parameters,  $F_s$ ,  $F_b$ ,  $R_s$ ,  $R_b$  is non-trivial. We identified them with the “pattern match” function in the Matlab optimization toolbox with a maximum of  $5 \times 10^6$  iterations. The optimal F and R parameters values for the shoulder joint region from the pattern search stage were identified as those producing the least amount of error across optimization intensities compared to the criterion intensity–endurance time relationships. We define the upper and lower bound of the rate parameters as  $F_{s_{lb}} = R_{s_{lb}} = -1.0$ ,  $F_{s_{ub}} = R_{s_{ub}} = 1.0$ ,  $F_{b_{lb}} = R_{b_{lb}} = 0.001$ ,  $R_{b_{ub}} = 0.0182 * 100$ , and  $F_{b_{ub}} = 0.00168 * 100$  following the conditions used for the TCM model (Jang et al., 2017). When the effect of  $BE$  is minimal (e.g.,  $F_s = 0$ ,  $R_s = 0$ ), we assume that our proposed LIN model behaves similar to the TCM model. Thus, in defining the initial rate parameters, we first compute the optimal parameters ( $F_{TCM}$ ,  $R_{TCM}$ ) of the TCM model. Then, the initial parameters of the LIN model were set as:  $F_{s,0} = R_{s,0} = 0$ ,  $F_{b,0} = F_{TCM}$ , and  $R_{b,0} = R_{TCM}$ .

### 3.2. Max shoulder torque estimation based on arm postures

In computing the shoulder torque (T), we used the biomechanical arm analysis implementation<sup>2</sup> provided by Jang et al. (2017). Based on the torque measurements (averaged torque) and the isometric load endurance task (endurance time), we could estimate each individual’s maximum shoulder torque (Jang et al., 2017). However, this method cannot capture the variance of the maximum torque based on the arm posture (Coury et al., 1998). To address this issue, we adopted

Chaffin’s model of arm strength (Chaffin et al., 1999). Chaffin’s strength model improves the estimation performance of shoulder torque. We will explain this further in the discussion section. In this model, the maximum shoulder torque value is estimated based on the shoulder and elbow angles as well as a biological sex parameter:

$$T_{max} = (227.338 + 0.525\alpha_e - 0.296\alpha_s) \times G_{shld}$$

where  $\alpha_s$  = shoulder flex angle,  $\alpha_e$  =  $180^\circ$  - elbow flex angle. We used  $\alpha_s = 90^\circ$ , and  $\alpha_e = 180^\circ$ , which describe the arm posture used in the isometric load task (see below). The biological sex adjustment parameter is defined as  $G_{shld} = 0.1495$  (female), and 0.2845 (male).

## 4. Methods

### 4.1. Participants

Twenty six right-handed participants (18 males;  $25.5 \pm 4.58$  yrs.;  $71.8 \pm 2.27$  kg; hand length  $20.3 \pm 2.26$  cm; lower arm length =  $27.4 \pm 3.28$  cm; upper arm length =  $35.6 \pm 2.66$  cm) volunteered in three experiments. Participants reported no musculoskeletal disorder or neurological disease. There were three experiments and they had a total of 20, 12, and 6 participants, respectively. The 12 participants in the second experiment were recruited from the first experiment. All participants provided written informed consent approved by the Institutional Review Board of Purdue University.

### 4.2. Equipment

We used a Microsoft Kinect sensor v2 with its corresponding software development kit (SDK) to track arm kinematics. We obtained joint torques from the camera data using inverse kinematic computations (Jang et al., 2017). We also used a moving-average filter (15th order) to smooth the joint-torque trajectories. Data was sampled at 50 Hz using a desktop computer with a Core i7 4.00 GHz CPU, 64 GB RAM, and an NVIDIA GTX 1080 GPU. A projector displayed the user interface on a  $1.6 \times 0.95$  meter screen placed 3 meters in front of the participant, at roughly the participants’ eye level. The Kinect camera was located between the participant and the screen; it was 1 meter in front of the screen and 1 meter above the floor, so that it did not interfere with the participant’s view of the screen.

Participants interacted with virtual targets projected on the screen using the Oculus Touch (Anon, 2020d) motion tracked controllers. A virtual pointer (green dot in Fig. 1a), mapped to the position of the controller held in the participant’s dominant right hand, was used to either point to targets, or virtually grasp and move objects projected on the screen. Vertical and horizontal movement of the pointer were controlled by vertical and medial–lateral movements of the controller, respectively.

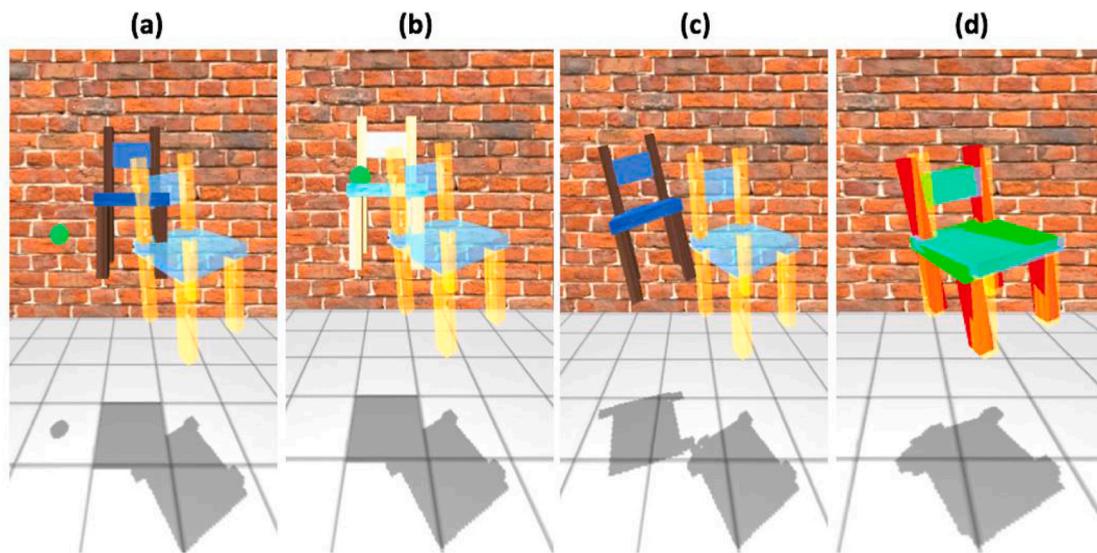
### 4.3. Experimental tasks

Participants first performed an isometric load task, and then performed one of three types of tasks.

#### 4.3.1. Isometric load task

To measure the maximum shoulder torque of each participant, we followed the indirect measurement method proposed by Jang et al. (2017). In this task, participants flexed the right shoulder to about  $90^\circ$  relative to the frontal plane and held a weight in their hand (2.27 kg for males and 1.36 kg for females) till volitional failure. Participants used visual feedback on the current and desired arm positions to maintain the required horizontal position of the arm.

<sup>2</sup> [https://github.com/CDesignGitHub/Cumulative-Arm-Fatigue\\_CHI-2017](https://github.com/CDesignGitHub/Cumulative-Arm-Fatigue_CHI-2017).



**Fig. 2.** Docking task interface. (a): “cursor” chair in the center of room (not clutched) and the “target” object to the right, (b): 3D pointer (green dot) “touches” the cursor chair, (c): the cursor chair is clutched to the pointer, (d): the cursor chair is docked to the target.

#### 4.3.2. Mid-air pointing task

We adopted the mid-air pointing task used by Jang et al. (2017). To start each trial, the participant stood in front of the screen with their right shoulder flexed in front of them at  $90^\circ$ , and the elbow fully extended, and with the Oculus controller in their hand. On the screen, the participant saw one circular target (10 cm diameter) and a smaller green cursor (4.8 cm diameter) which indicated the current position of the Oculus controller. The participant was instructed to control the position of the green cursor by altering their shoulder joint angles, while maintaining full elbow extension. The resulting vertical and horizontal movement of the hand-held controller was mapped to the movement of the cursor on the screen.

There were nine targets, arranged along the circumference of a circle (30 cm radius; Fig. 1b). When the participant’s shoulder was flexed at  $90^\circ$ , the cursor appeared at the center of this circle. During the pointing trials, one target appeared on the screen at a time. The participant rotated their shoulder to place the cursor into the target, and then pressed a button on the controller to indicate that they had matched the presented target. The current target then disappeared and another target appeared. The participant rotated their shoulder to place the cursor within the new target. This sequence continued for a pre-determined duration. The sequence of target presentation was pseudo-randomized such that the index of difficulty of each movement was consistent at 2.18. The participant matched all the presented target without lowering their arm between trials, and maintained the elbow at full extension throughout.

#### 4.3.3. Mid-air docking task

To validate our model in a more complex task involving translation and rotational movements of the arm and the object, participants were given a mid-air docking task in a rich virtual environment (Fig. 2). The environment minimizes ambiguity in the orientation by using a chair object instead of circles as the target and the cursor.

In each trial, a “cursor” chair appeared at the center of the room with neutral orientation (Fig. 2), and simultaneously, a “target” chair appeared at a different location and orientation. The participant moved their arm so that the controller (green dot in Fig. 2a) aligned with the cursor, and then pressed a trigger on the controller to “grab” the cursor chair. Participants then moved the cursor chair to a target chair through a combination of rotations and translations and aligned or docked the cursor with the target. Each trial was completed after the participant successfully aligned the cursor chair with the target

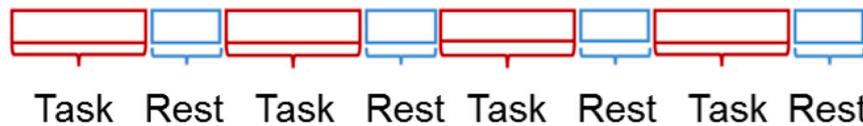
chair within a threshold for rotation ( $5^\circ$ ) and translation (3 cm). We provided auditory (beep sound) and visual (color change of target chair) feedback to indicate a successful docking, at which point, the participant released the trigger on the controller, signifying the end of the trial. Then, the current cursor and target chairs disappeared; a new cursor chair appeared at the same location as before, and a new target chair appeared at a different location and orientation. The participant moved their arm to grab the next cursor chair, without lowering the arm, and proceeded to dock the cursor with the new target. The target chairs were displayed at a randomly generated pose, but had the same amount of distance (30 cm) and rotation ( $45^\circ$ ) from the starting pose of the cursor chair. The starting position and orientation of the cursor chair was identical for all trials. We used our pilot data to determine the upper bounds of the range of motion of the target chair (i.e., 30 cm and  $45^\circ$ ) such that participants experienced fatigue, but were not so exhausted as to be unable to complete the experiment. Participants were instructed to maintain the elbow at full extension throughout this task.

This task was performed in discrete blocks, separated by rest intervals. The durations of the task and rest blocks were pre-defined, and they were altered across the three experiments (see below). For any task block, the timer began when the pointer came into contact with the first cursor chair, and the rest block began immediately after the end of the preceding task block. The total time of the docking task for experiments 2 and 3 was 685 s for each set, including both rest and task periods. Our pilot study confirmed that the participants found the docking task easy to learn and uncomplicated after an appropriate time of practice (i.e., 5 min).

#### 4.4. Procedure

We first measured body segment parameters, including total weight, height, upper arm (shoulder to elbow), lower arm (elbow to wrist), and hand (wrist to middle finger tip) lengths. The measurements were used to obtain the inertial properties of each segment (De Leva, 1996), which were then used to estimate shoulder torques via inverse dynamics (Jang et al., 2017).

Each participant started with a practice session and then performed the isometric load task, followed by both the pointing and docking tasks in experiment 1, or, in experiment 2 and 3, only the docking task.



**Fig. 3.** Randomized experimental protocol of pointing/docking task. For the shorter duration experiment, a task block was one minute, followed by either one of 5/10/15/20 s of rest. For longer durations, an experiment block consisted of either one of 100/120/140/160 s of task, followed by either one of 10/15/20/120 s of rest.

**Table 1**  
Borg CR10 scales with verbal anchoring.

Score	Definition	Note
0	Nothing At All	No arm fatigue
0.5	Very, Very Weak	Just noticeable
1	Very Weak	As taking a short walk
2	Weak	Light
3	Moderate	Somewhat but Not Hard to Go on
4	Somewhat Heavy	
5	Heavy	Tiring, Not Terribly Hard to Go on
6		
7	Very Strong	Strenuous. Really Push Hard to Go on
8		
9		
10	Extremely Strong	Extremely strenuous. Worst ever experienced

#### 4.4.1. Estimating maximum shoulder torque

For the isometric load task, we recorded the endurance time as the elapsed time from the beginning of the task till volitional failure, when the participant could no longer hold the weight in their outstretched dominant arm. The participants could see the Borg CR-10 scale (Table 1) at the side of screen, so that they could refer to the rating scale during the task. Participants were asked to report their subjective fatigue level using Borg ratings every 20 s. This task provides a good first-hand experience to participants in using Borg ratings (Jang et al., 2017), which facilitates its use in the subsequent tasks. The isometric task was followed by a mandatory 30-minute rest period to ensure sufficient recovery and minimal effect of this task on the subsequent tasks. Then, participants proceeded to either one of three experiments.

#### 4.4.2. Experiment 1: Comparing mid-air pointing and docking tasks

The purpose of experiment 1 was to validate the fatigue modeling approach of Jang et al. (2017) for more complex 3D interaction tasks, such as 3D docking. The other goal was to evaluate the generalizability of the model-based approach across different task types. For these reasons, this experiment consisted of two 3D interaction tasks: mid-air pointing and docking.

Participants had five minutes of mandatory practice for each task, or continued practice until they exhibited confidence in performing the tasks. The experimenter instructed participants to hit/dock as many as targets as possible. To follow the guidelines of Fitts' Law studies (Soukoreff and MacKenzie, 2004), instructions to strive for optimal performance, i.e. to 'hit as many targets as possible while staying accurate', contributed to keep each participant motivated. Once the participant was familiar with the tasks, they took a 15-minute mandatory break prior to the main tasks. Our pilot studies confirmed that these practice and rest duration were sufficient for minimizing potential learning effects and after-practice fatigue inference.

Subsequently, participants performed the pointing and docking tasks. We randomized the order of the tasks across participants. Each task had four one-minute blocks during which the trials were administered, and each task block was followed by either one of 5, 10, 15 or 20 s of rest (Fig. 3). Throughout the task, participants provided their perceived Borg ratings every 20 s. Participants were given a 15-minute rest period between the pointing and docking tasks to ensure adequate recovery.

#### 4.4.3. Experiment 2: Mid-air docking for varying task and rest durations

The purpose of experiment 2 was to validate the model with various durations of tasks and rests. Instead of a monotonous increase of fatigue level (experiment 1), we intended to evaluate the model performance when both increases and decreases of fatigue appear in the Borg recordings of participants.

We recruited 12 participants from experiment 1 on a separate day, more than a week afterwards, so that we could minimize any confounding factors from the prior experiment. Participants had the same training session as in experiment 1, followed by two sessions of mid-air docking tasks. Each session had randomly alternating task blocks of 100, 120, 140 and 160 s. Each task block was followed by either one of 10, 15, 20 and 120 s rest.

#### 4.4.4. Experiment 3: Individualized fatigue modeling

In our third experiment, similar to experiment 2, we evaluated our model with longer periods of task and rests. However, to validate that our model was also accurate for multiple recordings of each participant, participants performed mid-air docking tasks over a 7-day period. In this experiment, each participant performed the isometric task daily, followed by two sets of the docking task with alternating task and rest periods.

## 5. Results

In this section, we first show the variations of the Borg ratings in the experiments. Second, we present multiple cross-validations of interaction types and duration. We also evaluate our model-based approach across multiple trials from single users. Lastly, we draw comparisons between our new modeling method and an existing one (Jang et al., 2017) in each experiment.

### 5.1. Borg CR10 scale ratings across participants

Fig. 4(a) and (b) show the average Borg ratings of 20 and 12 participants during experiment 1 and 2, respectively. Although there was variability in the ratings across participants, the trend is similar across tasks. The rating increases linearly while performing the tasks and decreases linearly during rest periods, particularly during the longest rest period (120 s) where the rating decreases close to a 0 rating. We measured Borg ratings before and after each rest period. The average Borg rating for all participants in experiment 1 was 3.23 (SD = 1.28, Range = [0,8]), while the average rating for overlapping participants of experiments 1 and 2 was very similar, 3.45 (SD = 1.16, Range = [0,7]). Fig. 4(c) shows the variation of the Borg ratings in Experiment 3. This experiment used multiple recordings on the same participants (N = 6), with an average rating of 2.90 (SD = 0.90, Range = [0,8]).

### 5.2. Cross-validation of pointing and docking tasks

To investigate the applicability of the model for complex interaction tasks (i.e., beyond a simple pointing task), we performed a cross-validation test where the model is optimized with either the pointing or docking task. Then, we tested the model on the data set of the other task condition (Pointing  $\leftrightarrow$  Docking). Fig. 5 (Left) shows the cross-validation results over all participant data in experiment 1. The mean of root-mean-squared error (RMSE) during pointing tasks (Docking  $\rightarrow$  Pointing) was 1.43 (SD = 0.644), while the mean RMSE

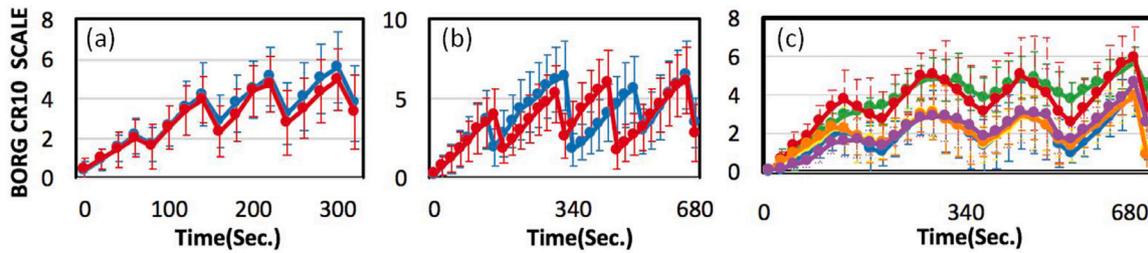


Fig. 4. Borg ratings of participants in all experiments. (a): Borg ratings by 20 participants during mid-air pointing (blue line) and docking (red line) tasks in experiment 1 (shorter task durations), (b): Borg ratings by 12 participants during mid-air docking tasks (blue: set 1, red: set 2) in experiment 2 (longer task durations), (c): Borg CR10 ratings by 6 participants during mid-air docking tasks for longer task duration in experiment 3. Each participant (indicated with different colors) repeated the experiment across 7 days.

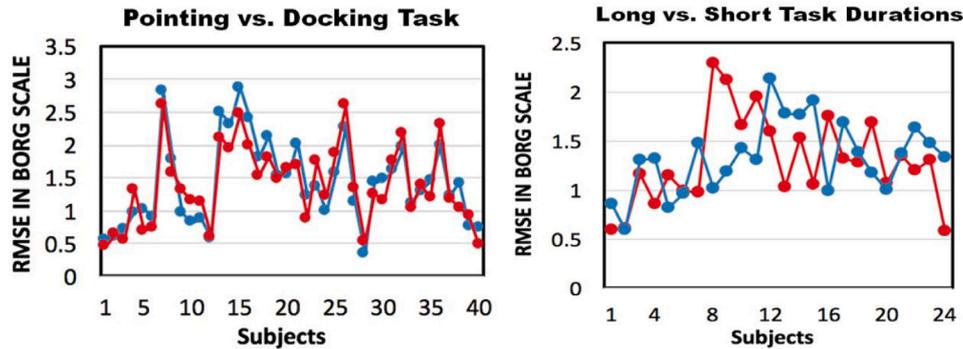


Fig. 5. *Left*: Results of the cross-validation between pointing and docking tasks. Blue line indicates the RMSE of the model optimized using the pointing task data set. Red line indicates the RMSE of the model optimized using the docking task data set. *Right*: Results of the cross-validation between shorter (Exp.1) and longer (Exp.2) durations of task/rest. Red line indicates the RMSE of the model optimized using experiment 1 data. Blue line indicates the RMSE of model optimized using experiment 2 data.

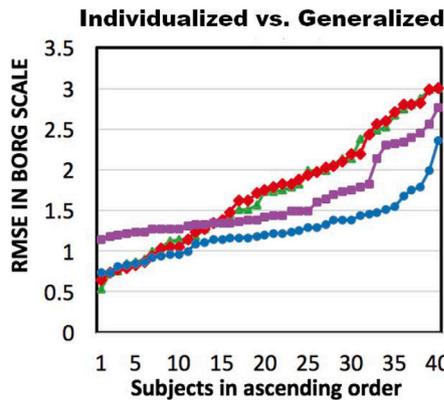


Fig. 6. Results of the leave-one-out cross validation of models using all participant data (generalized model) and the data of each participant (individualized model) in terms of RMSE for experiment 3. Green triangles: generalized TCM model, Red diamonds: generalized LIN, Purple squares: individualized TCM model, Blue circles: individualized LIN model.

during docking tasks (Pointing → Docking) was 1.39 (SD = 0.598). A paired-sample t-test showed no significant difference between Pointing and Docking tasks ( $t(39) = 0.932, p = 0.357$ ). Thus, we conclude that the complex task condition, i.e., docking task involving pointing, translating, and rotating, does not affect the estimation performance of the model.

### 5.3. Cross-validation of longer and shorter task periods

To evaluate the validity of the model for different interaction durations, we performed a cross-validation between shorter (experiment 1) and longer (experiment 2) durations of docking task. Fig. 5 (Right) shows the evaluation of performance between Experiments 1 and 2.

Mean RMSE in experiment 1 (experiment 2 → experiment 1) was 1.30 (SD = 0.45) and mean RMSE during experiment 2 (experiment 1 → experiment 2) was 1.33 (SD = 0.36). A paired-sample t-test showed no significant difference in model estimation performance between experiment 1 and 2 ( $t(23) = -0.28, p = 0.782$ ). Thus, we conclude that the duration of intermittent task and rest does not affect the estimation performance of the model.

### 5.4. Evaluating generalized and individualized models

To investigate the estimation performance of the model in a longitudinal experiment with single participants, we compared estimation performance between the individualized and generalized models. To assess the predictive performance of our model, we used a leave-one-out (LOO) cross-validation. With this, one dataset is excluded from optimizing the model, and then, that previously excluded dataset is used to test the model. Then this procedure is repeated in turn for every other dataset. We assume that the RMSE measured for each model is independent from each other when a single dataset is eliminated from the optimization. We obtained generalized models ( $M_{gen}$ ) from the entire data collected from experiment 3, while we generated individual models using the data set recorded from each participant ( $M_{ind}$ ). Results showed a RMSE of  $M_{gen} = 2.0$  (SD = 0.586) and  $M_{ind} = 1.25$  (SD = 0.315). A paired sample t-test showed a significant difference between the individualized and the generalized model ( $t(79) = 6.87, p = 0.000$ ). Fig. 6 further shows a comparison of our new model and the existing method (TCM (Jang et al., 2017)) for both individualized and generalized optimization. Our new modeling method (LIN) with individualized optimization showed consistently improved estimation accuracy. We conclude that our individualized model shows better estimation performance than a generalized model.

### 5.5. Comparison of modeling methods

To compare the estimation performance of our model with the existing one (the TCM-based model (Jang et al., 2017)), we used four

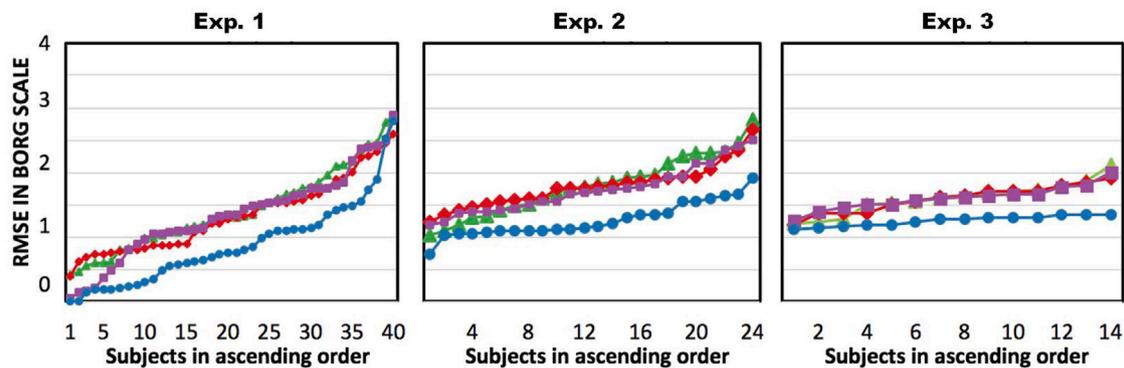


Fig. 7. Comparisons of the leave-one-out cross-validation results of the models in each experiment (experiment 1, 2 & 3). RMSE of estimation performance of four modeling methods are presented. Green triangles: TCM model w/o Chaffin's strength model, Red diamonds: LIN model w/o Chaffin's model, Purple squares: TCM w/ Chaffin's model, Blue circles: LIN model w/ Chaffin's model.

Table 2

Comparison of the estimation performance of four different fatigue quantification methods, for all investigated combination of fatigue models (LIN vs. TCM, with or without Chaffin's model). The estimation errors (RMSE) are measured in Borg ratings (0~10).

	Experiment 1	Experiment 2	Experiment 3
TCM w/o Chaffin	1.42	1.90	1.60
Lin w/o Chaffin	1.33	1.78	1.59
TCM w/ Chaffin	1.34	1.73	1.60
LIN w/ Chaffin	0.88	1.26	1.25

modeling methods to estimate the subjective fatigue in each experiment. The four models include (1) TCM without Chaffin's model, (2) TCM with Chaffin's model, (3) LIN without Chaffin's model, and (4) LIN with Chaffin's model (ours). These four models were separately optimized using the dataset for each experiment, experiment 1, 2 and 3. We followed the leave-one-out (LOO) cross-validation approach described in the prior section to test the robustness of each model in estimating fatigue of an unseen data set. In particular, in experiment 3, we performed LOO cross-validation within each participant's dataset to compare the model performance for estimating individual fatigue data in a longitudinal task period. In Table 2, the average RMSE of the Borg ratings for each fatigue modeling method is shown. Fig. 7 shows each model's estimation performance in terms of RMSE for each experiment. Overall, the LIN fatigue model combined with Chaffin's arm strength model consistently showed the best performance relative to the other methods. To further investigate the statistical difference among the RMSE results of the four modeling approaches, we performed a one-way ANOVA with a Tukey post-hoc test on the RMSE errors from the LOO cross-validation for each experiment. Results showed that only LIN with Chaffin's strength model has a statistically significant difference relative to the other three modeling methods ( $p < 0.0005$ ). The other three methods are not significantly different from each other ( $p > 0.8$ ).

### 5.5.1. Model parameters

The optimal parameters of our approach (LIN with Chaffin's model) are  $F_s = -0.043$  (SD = 0.011),  $F_b = 0.040$  (SD = 0.0035),  $R_s = 0.0$  (SD = 0.0),  $R_b = 0.0046$  (SD = 0.00097) in experiment 1; and  $F_s = -0.040$  (SD = 0.0075),  $F_b = 0.035$  (SD = 0.0015),  $R_s = 0.0$  (SD = 0.0),  $R_b = 0.0046$  (SD = 0.00040) in experiment 2. From the LOO cross-validation within each participant's data in experiment 3, we found the best performance model parameters and the overall range as  $F_s = 0.15$  (Range = [-0.045, 0.25]),  $F_b = 0.018$  (Range = [0.0010, 0.055]),  $R_s = -0.031$  (Range = [-0.051, 0.0]),  $R_b = 0.016$  (Range = [0.0052, 0.22]). Experiments 1 and 2 yielded similar optimal model parameters and they vary relatively less than the parameters from experiment 3. Also, as  $R_s$  is zero, the effect of Brain Effort (BE) on the rest rate was minimal. In contrast,

the optimal parameters varied in experiment 3 across participants. This indicates that the inter-individual difference in fatigue and rest rates are well reflected in the optimal parameters through the individualized modeling approach.

## 6. Discussion

The purpose of our setup that uses a Kinect to quantify subjective fatigue was to reduce the necessity for expensive or invasive equipment, such as dynamometers or EMG recordings. With our setup, our results showed the validity of our new model for quantifying cumulative subjective fatigue in various interaction conditions, including complex (i.e., mid-air docking task) and dynamic ones (i.e., varied rest/task periods).

### 6.1. Generalizability of the model

Our multiple cross-validations showed that the estimation performance of our model is not affected by task types (simple vs. complex tasks) nor periods (shorter vs. longer task and/or rest). These cross-validations simulated challenging estimation tasks where the model is optimized using a dataset from either one of two different task conditions, and then tested on the unseen data. The results imply the generalizability of our model to varying interaction conditions. For example, we can generate our model from an exemplar task condition and then use the model in other interaction conditions without having to expect a severe degradation of model performance.

### 6.2. Improved performance in estimating subjective fatigue

In comparing our new fatigue model (LIN) with an existing one (TCM (Jang et al., 2017)), we observed that the LIN model alone does not show a statistical improvement over the TCM model. Yet, we go beyond a simple addition of Chaffin's model to estimate maximum shoulder torque. When our LIN model is combined with Chaffin's strength model (Chaffin et al., 1999), we see that the estimation performance is significantly improved, with 31% less error on average in all experiments. In Fig. 4, we showed the large variations in the Borg ratings across participants. Given the variability and uncertain of subjective fatigue, estimating its accumulation during mid-air interaction is a challenging task. More specifically, it is challenging to predict how users *subjectively* experience their fatigue level during mid-air interaction. Thus, our improvement is an important stepping stone to accurately quantifying subjective user experience during mid-air interaction.

### 6.3. Individualized fatigue modeling

Our modeling approach showed an additional 42.5% reduction in fatigue estimation error when the longitudinal experiment data in experiment 3 is used for an individual participant's fatigue quantification rather than for all participants'. Individualized modeling accounts for the Borg ratings indicated by each participant, and thus it eliminates the variability in the fatigue perception across participants. We envision that the subjective fatigue ratings could be periodically captured while users perform mid-air interaction. Then, such information could be used to further inform the model as the capture of multiple ratings across time would reduce the variability in the model. This will further enhance the performance for estimating an individual's subjective fatigue at a given time.

### 6.4. Model parameters

Although we did not expect the negative optimal value of  $F_s$  for Exp. 1 and Exp. 2, this may have occurred to deal with the challenging optimization problem for a general model over 20 and 12 participants' subjective fatigue in Exp. 1 and 2, respectively. For instance, the  $BE$  ranged between 0 and 0.638, which defines the range of  $F$  in between 0.0142 and 0.040. Thus,  $F$  always shows a higher rate compared to  $R$  (0.0046). This means that increased  $BE$  still contributes to a higher rate of fatigue than recovery rate. In Exp. 3, we observed more understandable parameter optimization results where positive  $F_s$  and negative  $R_s$  values are achieved. We showed that inter-individual differences in fatigue and rest rates are well reflected in the optimal parameters through the individualized modeling approach.

From these observations, we may conclude that constructing a generalized subjective fatigue model from non-contact information is quite challenging and may not lead to understandable model optimization results (although the accuracy appears to be slightly enhanced as shown in Table 2). The best way to take advantage of our model is to parameterize individualized subjective fatigue as demonstrated in Exp. 3, which shows about 42.5% improvement over a single generalized model (see Figs. 6 and 7 (Right)).

Then, the optimal parameters for Exp. 3 can be also used to explain how our approach is different from a modified TCM model. For instance, Looft et al. provided modifications to the TCM model that substantially improved model predictions when intermittent tasks were involved (Looft et al., 2018; Looft and Frey-Law, 2020). However, when  $BE$  is 0, our model behaves similarly to this modified TCM as  $R$  becomes the max. When  $BE$  is not zero, the modified TCM becomes identical to the original TCM model. However, our approach features variable recovery rating  $R$  with respect to  $BE$ . This makes our modeling method superior to the original TCM (see Table 2 and Fig. 6) and may conceptually surpass the modified TCM. That said, the original TCM model is superior to other approaches when not accounting for the angle and velocity relationships on strength. In future work, we will generalize other modified models to handle our use case.

### 6.5. Simple and effective personalization in HCI scenarios

Our results showed an improvement in estimating an individual's subjective fatigue using only a camera-based skeleton tracking system instead of invasive and expensive tools that are impractical within HCI. In Kinect-based mid-air applications which can be adapted to each user's strength (e.g., exergames, at-home therapy), designers may enable users to calibrate the system with their own optimal ranges of motion, resting times, input positions, which involve significant physiological and psychological factors. Designers can also fit this approach to more complex scenarios by fitting the model parameters. Designers can run the model in any Kinect-based system as long as they collect a user's joint torque and Borg ratings corresponding data. Additionally, quantifying subjective fatigue could be used to develop

guidelines for collaborative work involving human-robot interaction, in which human exhaustion can be relieved by a robot. For instance, our model could be used to design industrial ergonomics and systems in order to protect industrial workers' health. Similarly, as technologies such as VR/AR as introduced in industry for upskilling and training purposes, our model can be useful to investigate human motion and fatigue in various training scenarios.

## 7. Limitations

Our model demonstrated promising results and estimation performance; however, we identified some limitations that go beyond the scope of this work, but that could be addressed in future work in this area. For example, we recognize that while our model measures elbow and shoulder torque, we have not validated the applicability of our model with varying grasping forces (i.e., the user holding a heavy object). While the weight of a controller (~0.169 kg) was a factor considered in the model, we have not validated our model for heavier objects, which will need higher grasping force. We also acknowledge that our participant pool is limited. Another limitation is that task difficulty and movement size (translation and rotation) were relatively limited. These bounds were established based on our pilot studies to avoid extreme exhaustion from our participants, so they could finish the experiment. To further validate the model, more challenging and dynamics tasks could be designed and tested.

## 8. Future work

Our model enables a variety of applications, also because the end effector (i.e., the controller weight) can be added as a factor when estimating shoulder torque into the model. An interesting direction will be to validate our model using different weights in the hand. Likewise, we plan to test on a participant pool with larger variability.

We also plan to validate our model for more dynamic and larger-range motions in the future. Another important direction will be to implement our model within real world applications.

Our future work might also include using our proposed model to quantify fatigue in other areas of the human body, particularly the upper body. Another interesting future direction is investigating and improving the efficiency of the heuristic pattern search for the discontinuous functions during our parameter optimization. We also plan to investigate the relationship between task accuracy and fatigue perception. Finally, we will release our model to the public so that it can be further validated and improved upon.

## 9. Conclusion

We presented an improved fatigue modeling method that incorporates the linear relationship between fatiguing/resting rates and brain effort (BE). We also incorporated a maximum arm strength model (Chaffin et al., 1999) into the model. Our new model showed significantly better estimation performance than previous work (Jang et al., 2017). Statistical analyses revealed that our model performance is not affected by conditions such as complexity of tasks and rest/interaction durations. Also, our results reveal that a personalized model can quantify an individual's subjective fatigue better than a general model. We investigated the relationship between the fatiguing/resting rate and the current muscular effort. Prior work had used constant rate parameters for this relationship (Jang et al., 2017). In the work presented here we showed that a linearly varying relationship significantly improved fatigue estimation performance. Our proposed method further generalizes and validates the applicability of the model-based fatigue estimation method.

We believe that improving and applying such a fatigue model for a variety of HCI scenarios, including mid-air tasks in AR and VR, will further validate the model-based approach, and will ultimately strengthen design guidelines targeted at minimizing fatigue during prolonged use of mid-air user interfaces.

## CRedit authorship contribution statement

**Ana Villanueva:** Conceptualization, Methodology, Data curation, Design, Writing – original draft. **Sujin Jang:** Conceptualization, Methodology, Software, Writing – original draft. **Wolfgang Stuerzlinger:** Conceptualization, Writing – review & editing. **Satyajit Ambike:** Conceptualization, Writing – review & editing. **Karthik Ramani:** Supervision, Conceptualization, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Amann, M., Romer, L.M., Pegelow, D.F., Jacques, A.J., Hess, C.J., Dempsey, J.A., 2006. Effects of arterial oxygen content on peripheral locomotor muscle fatigue. *J. Appl. Physiol.* 101 (1), 119–127.
- Anon, 2020a. Hololens 2. <https://www.microsoft.com/en-us/hololens>.
- Anon, 2020b. Oculus go. <https://www.oculus.com/go/>.
- Anon, 2020c. Oculus quest. <https://www.oculus.com/quest/>.
- Anon, 2020d. Oculus rift. <https://www.oculus.com/rift/>.
- Bachynskyi, M., Oulasvirta, A., Palmas, G., Weinkauff, T., 2014. Is motion capture-based biomechanical simulation valid for HCI studies?: study and implications. In: Proceedings of the 32nd Annual Conference on Human Factors in Computing Systems. ACM, pp. 3215–3224.
- Bachynskyi, M., Palmas, G., Oulasvirta, A., Steimle, J., Weinkauff, T., 2015. Performance and ergonomics of touch surfaces: A comparative study using biomechanical simulation. In: Proceedings of the 33rd Annual Conference on Human Factors in Computing Systems. ACM, pp. 1817–1826.
- Bigland-Ritchie, B., Furbush, F., Woods, J., 1986. Fatigue of intermittent submaximal voluntary contractions: central and peripheral factors. *J. Appl. Physiol.* 61 (2), 421–429.
- Bijur, P.E., Silver, W., Gallagher, E.J., 2001. Reliability of the visual analog scale for measurement of acute pain. *Acad. Emergency Med.* 8 (12), 1153–1157.
- Borg, G.A., 1982. Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc.* 14 (5), 377–381.
- Carifio, J., Perla, R.J., 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *J. Soc. Sci.* 3 (3), 106–116.
- Chaffin, D.B., Andersson, G., Martin, B.J., et al., 1999. Occupational biomechanics. Wiley New York.
- Cifrek, M., Medved, V., Tonković, S., Ostojić, S., 2009. Surface EMG based muscle fatigue evaluation in biomechanics. *Clin. Biomech.* 24 (4), 327–340.
- Coury, H.J.C.G., Kumar, S., Rodgher, S., Narayan, Y., 1998. Measurements of shoulder adduction strength in different postures. *Int. J. Ind. Ergon.* 22 (3), 195–206.
- De Leva, P., 1996. Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters. *J. Biomech.* 29 (9), 1223–1230.
- Enoka, R.M., Stuart, D.G., 1992. Neurobiology of muscle fatigue. *J. Appl. Physiol.* 72 (5), 1631–1648.
- Frey-Law, L.A., Looft, J.M., Heitsman, J., 2012. A three-compartment muscle fatigue model accurately predicts joint-specific maximum endurance times for sustained isometric tasks. *J. Biomech.* 45 (10), 1803–1808.
- Frey-Law, L.A., Schaffer, M., Urban III, F.K., 2021. Muscle fatigue modelling: Solving for fatigue and recovery parameter values using fewer maximum effort assessments. *Int. J. Ind. Ergon.* 82, 103104.
- Fuglevand, A.J., Winter, D.A., Patla, A.E., 1993. Models of recruitment and rate coding organization in motor-unit pools. *J. Neurophysiol.* 70 (6), 2470–2488.
- Gede, G., Hubbard, M., 2014. A bioenergetic model for simulating athletic performance of intermediate duration. *J. Biomech.* 47 (14), 3448–3453.
- 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.
- Hassenzahl, M., Tractinsky, N., 2006. User experience—a research agenda. *Behav. Inf. Technol.* 25 (2), 91–97.
- Hayes, K., Walton, J.R., Szomor, Z.L., Murrell, G.A., 2002. Reliability of 3 methods for assessing shoulder strength. *J. Shoulder Elbow Surgery* 11 (1), 33–39.
- Hincapié-Ramos, J.D., Guo, X., Moghadasian, P., Irani, P., 2014. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In: Proceedings of the 32nd Annual Conference on Human Factors in Computing Systems. ACM, pp. 1063–1072.
- Jang, S., Stuerzlinger, W., Ambike, S., Ramani, K., 2017. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. CHI '17, ACM, New York, NY, USA, pp. 3328–3339. <http://dx.doi.org/10.1145/3025453.3025523>, URL <http://doi.acm.org/10.1145/3025453.3025523>.
- Kinali, G., Kara, S., Yıldırım, M.S., 2016. Electromyographic analysis of an ergonomic risk factor: overhead work. *J. Phys. Therapy Sci.* 28 (6), 1924–1927.
- Law, L.A.F., Lee, J.E., McMullen, T.R., Xia, T., 2010. Relationships between maximum holding time and ratings of pain and exertion differ for static and dynamic tasks. *Applied Ergon.* 42 (1), 9–15.
- Looft, J.M., Frey-Law, L.A., 2020. Adapting a fatigue model for shoulder flexion fatigue: Enhancing recovery rate during intermittent rest intervals. *J. Biomech.* 106, 109762.
- Looft, J.M., Herkert, N., Frey-Law, L., 2018. Modification of a three-compartment muscle fatigue model to predict peak torque decline during intermittent tasks. *J. Biomech.* 77, 16–25.
- Morishita, S., Yamauchi, S., Fujisawa, C., Domen, K., 2014. Rating of perceived exertion for quantification of the intensity of resistance exercise. *Int. J. Phys. Med. Rehabil.* 2013.
- Öberg, T., Sandjö, L., Kadefors, R., 1994. Subjective and objective evaluation of shoulder muscle fatigue. *Ergonomics* 37 (8), 1323–1333.
- Rashedi, E., Nussbaum, M.A., 2015. Mathematical models of localized muscle fatigue: sensitivity analysis and assessment of two occupationally-relevant models. *PLoS One* 10 (12), e0143872.
- Roy, J.-S., Ma, B., MacDermid, J.C., Woodhouse, L.J., 2011. Shoulder muscle endurance: the development of a standardized and reliable protocol. *BMC Sports Sci. Med. Rehabil.* 3 (1), 1.
- Segerstrom, S.C., Nes, L.S., 2007. Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychol. Sci.* 18 (3), 275–281.
- Sjøgaard, G., Savard, G., Juel, C., 1988. Muscle blood flow during isometric activity and its relation to muscle fatigue. *Euro. J. Appl. Physiol. Occupational Physiol.* 57 (3), 327–335.
- Sonne, M.W., Potvin, J.R., 2016. A modified version of the three-compartment model to predict fatigue during submaximal tasks with complex force-time histories. *Ergonomics* 59 (1), 85–98.
- Soukoreff, R.W., MacKenzie, I.S., 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in HCI. *Int. J. Hum.-Comput. Stud.* 61 (6), 751–789.
- Vuibert, V., Stuerzlinger, W., Cooperstock, J.R., 2015. Evaluation of docking task performance using mid-air interaction techniques. In: Proceedings of the 3rd ACM Symposium on Spatial User Interaction. ACM, pp. 44–52.
- Xia, T., Law, L.A.F., 2008. A theoretical approach for modeling peripheral muscle fatigue and recovery. *J. Biomech.* 41 (14), 3046–3052.



**Ana Villanueva** is a Ph.D. candidate in Human-Computer Interaction (HCI). She received her undergraduate degree in Mechanical Engineering from the University of Kansas. Her current research sits at the intersection of HCI and the learning sciences. During her Ph.D. work in the C-Design Lab at Purdue University, she has worked on designing multiple prototypes and STEM K-12 learning content to provide high-quality educational tools. Her thesis is titled “Investigating New Modalities and Interaction Techniques of Multimedia in STEM Education”. Her work uses primarily augmented-reality applications to teach programming and electrical circuitry.



**Dr. Sujin Jang** received the B.S. in Mechanical and Automotive Engineering from Kookmin University, in 2010, the M.S. in Mechanical Engineering from the University of Florida, in 2012, and the Ph.D. in Mechanical Engineering from Purdue University, in 2012. His past and current research work broadly involve deep learning, human-computer interaction, visual analytics, and robotics.



**Wolfgang Stuerzlinger** Building on his deep expertise in Virtual Reality and Human-Computer Interaction, Dr. Stuerzlinger is a leading researcher in Three-dimensional User Interfaces. Since 2014, he is a full professor at the School of Interactive Arts + Technology at Simon Fraser University in Vancouver, Canada. Current research projects include better 3D interaction techniques for Virtual and Augmented Reality applications, new solutions for Visual and Immersive Analytics, the characterization of the effects of technology limitations on human performance, investigations of human behaviors with occasionally failing technologies, user interfaces for versions, scenarios, and alternatives, and new Virtual/Augmented Reality hardware and software.



Dr. **Satyajit Ambike** is a Mechanical Engineer, with a research focus on movement in biological systems. He currently focuses on human locomotion and prehension across the adult age span. He is interested in how aging influences stability of various movements, and how stability and maneuverability – two antagonistic movement attributes – are traded off in activities of daily living. He is currently studying synergistic control of adaptive locomotion, digit-object interactions, and the quantification of fatigue during human-computer interactions.



**Karthik Ramani** is an Indian born American researcher, mechanical engineer, and entrepreneur. He is the Donald W. Feddersen Distinguished Professor of Mechanical Engineering at Purdue University, with courtesy appointments in Electrical and Computer Engineering and Educational Studies in the College of Education. Previously, from 2001–2008, he was the Director of the Center for Information Sciences in Engineering at Purdue University. Ramani is best known for his work on computer vision, human-computer interaction, and computational fabrication.