

Modeling Cumulative Arm Fatigue in Mid-Air Interaction based on Perceived Exertion and Kinetics of Arm Motion

¹Sujin Jang, ²Wolfgang Stuerzlinger, ³Satyajit Ambike, ¹Karthik Ramani*

¹School of Mechanical Engineering and

²School of Interactive Arts + Technology

³Department of Health and Kinesiology
Purdue University, West Lafayette, IN, USA
{jang64, ramani, sambike}@purdue.edu

Simon Fraser University
Vancouver, CA
w.s@sfu.ca

ABSTRACT

Quantifying cumulative arm muscle fatigue is a critical factor in understanding, evaluating, and optimizing user experience during prolonged mid-air interaction. A reasonably accurate estimation of fatigue requires an estimate of an individual's strength. However, there is no easy-to-access method to measure individual strength to accommodate inter-individual differences. Furthermore, fatigue is influenced by both psychological and physiological factors, but no current HCI model provides good estimates of cumulative subjective fatigue. We present a new, simple method to estimate the maximum shoulder torque through a mid-air pointing task, which agrees with direct strength measurements. We then introduce a cumulative fatigue model informed by subjective and biomechanical measures. We evaluate the performance of the model in estimating cumulative subjective fatigue in mid-air interaction by performing multiple cross-validations and a comparison with an existing fatigue metric. Finally, we discuss the potential of our approach for real-time evaluation of subjective fatigue as well as future challenges.

Author Keywords

Mid-air interaction; perceived exertion; cumulative fatigue model; maximum arm strength; biomechanical arm model.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION

Recent advances in sensing technologies have produced low-cost human motion tracking methods (e.g., MS Kinect, Leap Motion), and these methods are being leveraged to study and adopt natural human motion for the design of mid-air interactions in various environments (e.g., AR/VR [37, 40], medical operations [36], gaming [33], and virtual design [54]). Physical ergonomics is an important design factor for mid-air interaction [8, 26, 30]. In particular, arm fatigue – the

so-called “Gorilla arm” effect [2] – is known to negatively impact user experience and hamper prolonged use of mid-air interfaces. Fatigue is a complex phenomenon defined as the decline in maximal force or power capacity of muscle due to sustained activation [20]. This means that sub-maximal forces, such as those encountered in most mid-air interactions, can be sustained for extended periods even as fatigue accumulates. However, users will typically perceive the accumulation of fatigue (i.e., feel the arm getting heavier), and this may negatively impact their experience. Furthermore, subjective fatigue and its perceived intensity are influenced not only by the task demands and the individual's current physiological state, but also by the history of the individual's movement choices. Consequently, models estimating the current fatigue of a user must be based on continuous measurement of the ongoing interaction. Although the term “Arm fatigue” appears in more than 300 publications in HCI over the last decade, we do not have good predictive models of cumulative fatigue that account for variable and intermittent periods of interaction.

Another practical issue in quantifying muscle fatigue is the measurement of an individual's maximal strength. The fatigue rate for a given task depends on the proportion of muscular strength that the task consumes. Naturally, fatigue estimation requires the quantification of an individual's maximal ability [19], and biomechanists typically achieve this by recording the maximal forces (or joint torques) that the subject can produce under isometric conditions [34, 41, 44]. However, these measurements require expensive force sensors and specialized test rigs, making sensor-based strength measurements impractical for HCI research.

Lastly, physiological and psychological human factors both influence fatigue [20], and thus, user experience as well [28]. To better evaluate and understand the effect of fatigue on user experience, we need congruent models that account for both influences [28, 32]. Current models to quantify fatigue are mostly informed by direct measurements of forces or physiological signals (e.g., movement dynamics, muscle forces), and rarely account for the perceived effect of fatigue accumulation. These direct measurements are invasive, require specialized equipment, and therefore, are not practical for implementation in HCI. Subjective fatigue evaluation (e.g., Borg scales [11]) requires no set-up and provides useful information of the user experience. However, direct verbalization and reporting is necessary, which, again interferes with the user's interactive experience. Thus, existing objective and subjective methods for evaluating fatigue can be cumbersome and invasive as multiple recordings are required during tasks.

*School of Electrical and Computer Engineering (by courtesy)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2017, May 6–11, 2017, Denver, CO, USA.

Copyright © 2017 ACM ISBN 978-1-4503-4655-9/17/05 ...\$15.00.

<http://dx.doi.org/10.1145/3025453.3025523>

In HCI, researchers recently investigated the role of muscle fatigue [7, 8, 30]. However, previous work mostly considers instantaneous muscle contractions, and thus cannot quantify cumulative muscle fatigue. Also, previous work is informed by physical measures (e.g., muscle forces and torques), and does not quantify fatigue perception, which is necessary for the accurate evaluation of user experience during interaction.

This paper addresses the above-mentioned issues in evaluating cumulative arm fatigue during mid-air interaction. First, we propose a simple and effective method to estimate an individual's maximal shoulder torque using a mid-air pointing task. We use a camera-based skeleton tracking system to capture arm movements, and then compute the shoulder-joint forces and torques during the pointing task, which subjects perform till volitional failure. Then, we use the *Intensity–Endurance Time* curve [41] to estimate the maximum shoulder torque without a direct force or torque measurement. *Intensity* is the ratio between the maximum and current torque exerted at the shoulder. *Endurance time* (ET) is the duration of muscular exertion before the volitional failure. We validate our indirect maximum torque estimation method via comparisons with direct measurements obtained with a force sensor.

Second, we introduce a method to quantify cumulative arm fatigue based on the three-compartment muscle model (TCM) [56]. The model allows muscle units to be in either an active (M_A), fatigue (M_F), or rest (M_R) state and defines state-transition rules. This enables quantification of cumulative fatigue when both task and rest periods appear during interaction. By optimizing the free model parameters to minimize the difference between the model and user-reported estimates of perceived fatigue, we account for both physical and psychological processes that influence perceived fatigue. We investigate the performance of the model in predicting subjective fatigue in mid-air pointing tasks through cross-validations and a comparison with an existing metric.

Our contributions include: (1) a simple and effective method to estimate the maximum shoulder torque through a mid-air pointing task; (2) a cumulative fatigue model that estimates subjective fatigue using movement kinetics measured during ongoing mid-air interaction; (3) experimental validation of the cumulative fatigue model for estimating subjective arm fatigue; and (4) a procedure for modeling the cumulative fatigue that accounts for subject specificity of muscle fatigue. Below, we describe related work, techniques, experimental methods, and the analysis of experimental results. Our findings show the potential of the model in predicting subjective fatigue in mid-air interactions. We discuss the challenges for further developing the model to quantify arm muscle fatigue derived from subjective and biomechanical factors.

RELATED WORK

Here we review prior work related to objective and subjective evaluation of muscle fatigue, cumulative fatigue models, and muscle fatigue evaluation for mid-air interactions.

Objective Fatigue Evaluation

Muscle fatigue is quantified based on changes in various physiological quantities such as muscle activation (measured

using electromyography (EMG) [14]), intra cellular acidosis [55], arterial oxygen [3], blood pressure [49], and heart rate [48]. Such invasive methods are impractical for HCI.

Another method uses a measurement of maximum muscular strength. The maximum capacity of muscular exertion is defined as the “*maximum voluntary contraction* (MVC)”, and it is commonly expressed as a force (N) or a torque ($N \cdot m$). Reduction in MVC during a continuous task is considered as a reliable assessment of muscle fatigue. Biomechanists commonly measure MVC using force sensors in specialized test rigs (e.g., [1, 29, 47]). However, such set-ups are invasive, not easily accessible, and impractical for evaluating muscle fatigue during HCI tasks.

Ratings of Perceived Exertion

There are several psychometric scales commonly used for evaluating subjective fatigue such as Likert scale [13], NASA-TLX [27], VAS (visual analog signal) [9], and ratings of perceived exertion (RPE) [12]. Such subjective evaluation based on the muscular sensation is also a result of physiological changes [4], and it should be carefully considered in evaluating muscle fatigue. The Borg RPE scales, given their reproducibility and sensitivity [12], have been extensively used in psychological evaluation of muscle fatigue as a complement to physiological measures in various field of research (e.g., sports [21], medicine [43], ergonomics [10]). Borg CR10 [12] is a categorical rating with ratio properties (0~10 points), and it has been used to assess perceived exertions in body parts such as the hand and the arm [22, 39, 51]. The Borg CR10 scale uses verbal anchors and numbers to map the magnitude of exertion to a scalar invariance scale.

Several studies have shown that the Borg CR10 scale is strongly correlated with objective measures, such as EMG [17, 53], and that it is reliable at lower levels of force exertion (e.g., holding a weight less than 2~3 kg [44]). This makes the Borg CR10 scale particularly suitable for evaluating fatigue in mid-air interaction, where relatively little or no external load is exerted on the arms (bare-hand interactions).

Cumulative Muscle Fatigue Model

Due to the analytic complexity of describing fatigue, few models are available to properly quantify muscle fatigue. The Endurance Time (ET)–Intensity curve (e.g., Rohmert's curve [46]) has been extensively used to empirically model the effect of fatigue on the task endurance time in static exertion tasks. Although this approach is useful in estimating fatigue-induced ET, it cannot be used for more complex tasks involving dynamic load conditions or recovery during rest periods. Liu et al. [38] proposed a dynamic muscle model based on three muscle states: rest (M_R), active (M_A), and fatigue (M_F). However, this model has only been validated under constant, maximum effort, and thus it is not generalizable to dynamic load conditions such as arm movements interspersed with rest periods. Xia et al. [56] proposed a three-compartment muscle (TCM) model by introducing a feedback controller term between the active (M_A) and rest (M_R) muscle states. More recently, Frey-Law et al. [24] validated the TCM model in estimating ET under static load conditions,

and obtained a set of joint-specific model parameters. This enables the model to predict the muscle activation states under various task conditions (i.e., static and dynamic load conditions). However, their validation was still limited to static load conditions. Sonne et al. [50] further experimentally validated the performance of the TCM model in predicting fatigue during grasping and thumb contraction tasks by providing a task-specific, optimized fatigue model.

Similarly, we optimize the TCM model for estimating shoulder joint fatigue. Also, we optimize the model using both subjective fatigue ratings and movement kinematics. Once optimized, our model estimates previously unseen subjective fatigue levels based only on kinematic measurements. We are not aware of any prior work that achieves this.

Quantifying Muscle Fatigue in Mid-Air Interactions

In HCI, there are two closely related studies on quantifying arm fatigue during mid-air interaction. Consumed Endurance (CE) [30] is an easy-to-implement metric for estimating ET in mid-air selection tasks. The authors showed a strong correlation between CE and the Borg scale. Yet, CE is based on Rohmert's ET curve [46], which is only valid for static load conditions, and thus CE does not consider rest and its effect on cumulative fatigue [56]. Also, CE is assumed to be zero at exertion levels below 15%. This controversial assumption [23, 41] limits use of CE for evaluating mid-air interaction with low exertion levels. CE also uses generic maximum shoulder torque values, even though shoulder strength is subject-specific [41]. In contrast, we propose a novel method to measure an individual's maximum shoulder torque that does not require expensive sensors, and which shows strong agreement with traditional, contact-based measurements.

Bachynsky et al. [7, 8] introduced a biomechanical simulation to evaluate muscle activation. This work showed the potential of simulation-based approaches to quantify the effect of fatigue on the design of touch-based and mid-air interactions. However, the simulation only considers muscle contractions in a static optimization, and thus cannot quantify cumulative muscle fatigue. Also, Bachynsky et al. [6] observed that their simulation output does not correlate well with subjective fatigue measures. This makes their simulation results less comparable with our model estimates.

Unlike existing approaches, we consider cumulative arm fatigue involving both interaction and rest periods without limiting the range of exertion. Furthermore, our model quantifies the user's perception of fatigue, which is likely more directly related to the user experience during mid-air interaction.

ESTIMATING ARM MUSCLE FATIGUE

Here we describe our method for quantifying arm fatigue by introducing three-compartment muscle model, biomechanical arm model, and maximum shoulder torque estimation.

Three-Compartment Fatigue Model

We introduce the three-compartment muscle (TCM) model for estimating active, rest, and fatigue muscle states (see [24, 56] for more details). The TCM model assumes that the motor units involved in a task can be in one of three possible

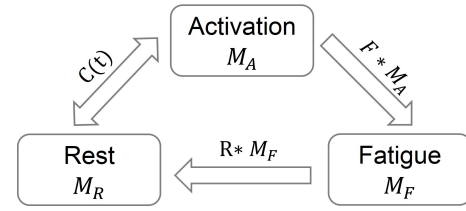


Figure 1. Three-compartment muscle (TCM) model

states: *active* (motor units receiving neural activation and contributing to the current task), *fatigued*, (fatigued motor units without activation), or *resting* (inactive motor units that are not required for the current task). Figure 1 shows the three muscle unit states, and the relations between them. $M_F(\%)$ is the proportion, or ‘compartment’ of motor units that are currently in fatigued state, $M_A(\%)$ is the compartment of active motor units, and $M_R(\%)$ is the proportion of motor units at rest. Each compartment of motor states is expressed as a percentage of the *maximum voluntary contraction* (%MVC). Since motor-unit recruitment is binary (they are either contracted or not), $M_A = 100\%$ indicates that all motor units are recruited for a MVC task. A sub-maximal task implies $M_A < 100\%$. The three motor unit states are related 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 , are the model parameters defining the rate at which active motor units are fatiguing, and the rate at which fatigued motor units recover and enter the rest state, respectively, and $C(t)$ is motor unit activation function defined as:

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

where TL is the target load defined as a torque ratio $[T_{current}/T_{max}] * 100(\%)$, L_D is the muscle force development factor, 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. [24].

Model Fitting

We used the optimization toolbox in MATLAB to optimize the TCM and to identify the model parameters (F and R) for mid-air interaction. The pattern search method [5] was used to minimize the root mean squared errors (RMSE) between the perceived exertion (Borg scale ratings) and the size of the fatigued motor unit compartment (M_F):

$$\begin{aligned}\underset{F, R}{\text{minimize}} \quad & \sqrt{\frac{1}{n} \sum_{i=1}^n [\phi(M_F(i)) - B(i)]^2} \\ \text{subject to} \quad & F \in [F_{lb}, F_{ub}], R \in [R_{lb}, R_{ub}]\end{aligned}$$

where n is the number of fitting data, $M_F(i)$ is the fatigue level estimation, $B(i)$ is the Borg scale rating, and the upper and lower bounds of the parameters are defined as $F_{lb} =$

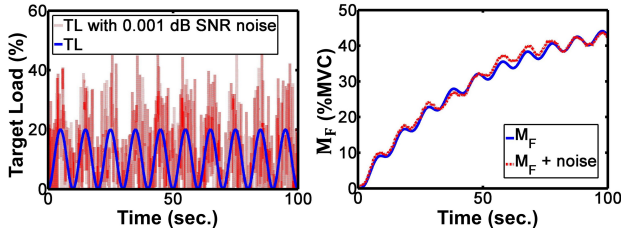


Figure 2. Simulated fatigue estimates (M_F) using the TCM model with 0.001 dB noise. Left: Input (TL) to the model w/ and w/o noise. Right: comparison of estimation results w/ and w/o noise.

$R_{lb} = 0.001$, $F_{ub} = 0.0182 * 100$, and $R_{ub} = 0.00168 * 100$. The upper bounds are defined by multiplying the constant parameters recommended by Frey-Law et al. [24] with 100. We assume a linear relationship between the Borg scale and %MVC (proportion of current and maximum torque) based on a review of their relationship [42]. $\phi(\cdot)$ is a linear function mapping the fatigue estimation $M_F(i)$ to the Borg scale. We define the linear mapping as $\phi(x) = 0.0875 * x$. This means a value of ‘8’ in Borg CR10 scale is equivalent to 90% of M_F .

Sensitivity Analysis of Fatigue Estimation

Since we target the use of off-the-shelf vision-based skeleton tracking systems (e.g., Kinect), we need to deal with the fact that such tracking is more susceptible to noise and jitters compared to marker-based tracking systems (e.g., Vicon, OptiTrack). Therefore, we tested the sensitivity of the TCM model to the presence of severe noise. Simulation results (see Figure 2) showed that the model is relatively insensitive to jitters and errors in torque estimation (mean absolute error, $M = 1.00$, $SD = 0.98$ in %MVC).

Biomechanical Upper Limb Model

Fatigue development rate is joint specific, and the shoulder fatigues much faster than the elbow or wrist during arm movements [23]. Therefore, we assume that arm fatigue is mostly attributable to shoulder-joint fatigue, and the shoulder-joint strength is the most relevant design-related parameter for mid-air interaction.

Estimating shoulder fatigue using the TCM model requires two inputs: the maximum joint torque (T_{max}) and the time-course of torque exerted at the joint ($T(t)$). Our new method to estimate T_{max} will be discussed in the next section. Here we briefly introduce a standard procedure to measure the shoulder joint torque $T(t)$ based on a biomechanical arm model [15]. We model the upper limb as rigid bodies (*links*) connected in series by *joints* [45]. Figure 3 shows the model defined by a 4 degree-of-freedom (DOF) serial chain (shoulder:3, elbow:1). The biomechanical arm model equations and the Newton-Euler inverse dynamics procedure [15] to compute the shoulder torque (T) are provided in the Appendix.

Estimating Maximum Shoulder Torque

To estimate the muscle strength of users, we propose a procedure where users perform a fatiguing exercise by holding a weight in the air. This procedure eliminates the need of a force sensor and dynamic measurements for estimating the

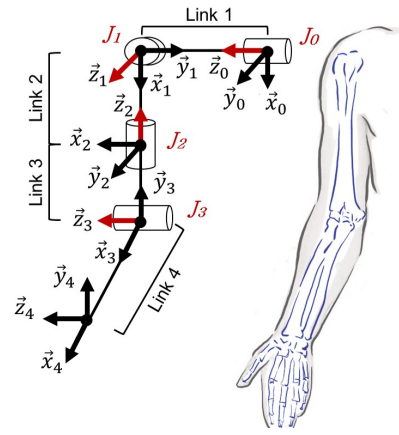


Figure 3. Biomechanical model of the upper limb.

maximal shoulder torque. In this isometric exercise, users are asked to flex the shoulder and maintain the arm in the horizontal position while holding a weight (e.g., 1.36 kg for female and 2.27 kg for male subjects). To guide subjects in maintaining this posture, a simple pointing task is designed. A circular target is fixed at the same height as the shoulder, and placed in front of the body at maximum arm extension (see Figure 4(b)). The target is displayed on screen, and can only be “touched” when the user holds the correct posture.

We compute the time course of the shoulder torque $T(t)$ using the upper limb biomechanics model and measure the endurance time while subjects perform the isometric exercise. The subjects are assumed to exert constant shoulder torque for this isometric test, and we compute the average shoulder torque exerted during the exercise (T_{avg}). We use the equation of maximal shoulder capacity (%MVC) and endurance time (ET) experimentally derived by Mathiassen et al. [41]: $ET = a * \exp(-b * \%MVC)$, where %MVC is the ratio of current and maximum shoulder torque ($T_{current}/T_{max} * 100\%$), and a and b are experimentally defined constants (40.609 and 0.097). To find the maximum torque T_{max} , we invert Mathiassen’s equation as:

$$T_{max} = \frac{-b * T_{avg}}{\log(ET/a)} \times 100 \quad (1)$$

This yields an estimate of the the maximum shoulder torque (involving the anterior deltoid, the coracobrachialis, biceps, and pectoralis major muscles). We compare this indirect measurement with traditional sensor-based ones in our experiments.

EXPERIMENTAL METHODS

With the two experiments described below, we validate (1) a method to estimate maximal shoulder-joint torque using kinematic measurements with the Kinect camera and (2) the three-compartment muscle (TCM) model for estimating cumulative shoulder fatigue during mid-air interaction. We focus on a “virtual hand” mid-air pointing task, a fundamental mode of interaction in 3D user interfaces.

General Setup

We used a Microsoft Kinect sensor and the corresponding SDK to track arm movements. Data were sampled at 50

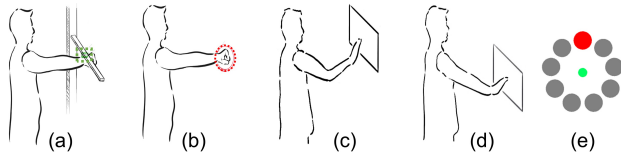


Figure 4. (a): sketch of the dynamometer setup, (b): posture for isometric exercise, (c) and (d): different interaction zones for pointing task, and (e): target circles arranged on a 2D plane. Target width is 10 cm and distance is 30 cm (red: target to select, green: pointer).

frames/second using a laptop with a Core i7 2.90GHz CPU and 16 GB RAM. The computer was connected to a short-throw projector to display the task interface. The physical size of projected screen on the wall was 1.6×0.95 meter, and the Kinect camera was located 1 meter in front of the screen and 1 meter above the floor. Participants stood 3 meters in front of the screen. The setup was identical for both experiments. Joint torques were obtained from the camera data using the inverse dynamics computations described in the Appendix. A moving-average filter (15^{th} order) was used to smooth the joint-torque trajectories.

Mid-Air Pointing Tasks

To systematically collect arm movements and analyze resultant arm fatigue, we designed a mid-air pointing experiment based on the ISO 9241-9 standard [35, 52], to minimize the impact of confounding factors on the subjective fatigue measurement during the pointing tasks. This task has participants point at a circle of targets with a given width in a sequence that matches the movement distance to the circle diameter (see Figure 4(e)). Haga et al. [25] noted that task difficulty significantly affects subjects' perceived fatigue. In our experiments, we aim to measure perceived fatigue level only affected by muscular effort, not by task difficulty. We observed during pilot studies that learning affects the perceived fatigue. Consequently, it is important to provide enough training time and ensure that the pointing task was not too challenging. Thus, we decided to fix the task difficulty by using a single index of difficulty (ID). We also noticed that the tracking performance of the Kinect can affect the ID of pointing tasks as it may interfere with movement time and error rate. In our pilot studies, we tested a range of target widths (5.0 to 15.0 cm) with a fixed distance (30 cm) to identify those that show no degradation of pointing performance in terms of throughput. In the end, we decided to use a target width of 10 cm width at 30 cm distance, corresponding to an ID of 2.18.

Participants

We recruited 24, right-handed volunteers (20 male, 4 female) from the local university campus. The age ranged from 20 to 34 years ($M = 25.5$ yrs); height ranged from 154 to 185 cm ($M = 174.1$ cm); weight from 48.1 to 112 kg ($M = 73.6$ kg); upper arm length from 25.5 to 38 cm ($M = 33.5$ cm); lower arm length from 22.5 to 29.5 cm ($M = 26.8$ cm); and hand length ranged from 17.5 to 22 cm ($M = 20$ cm). No subject had a musculoskeletal disorder or neurological disease. The average duration of experience with mid-air interfaces ranged from 0 to 5 years ($M = 0.6$ yrs).

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

Table 1. Borg CR10 scales with verbal anchoring.

Design

Each subject first performed a training session. Subsequently, he/she performed two experimental tasks: (1) maximum shoulder torque generation and (2) mid-air pointing.

Training

Pilot testing confirmed that subjects learned the pointing tasks within a few (< 5) minutes, and that learning affected subjective fatigue ratings. Therefore, to minimize any potential impact of learning on the perceived exertion, we provided at least 5 minutes of training where subjects practiced the experimental pointing task (Exp. 2). After the training session, we provided at least 5 minutes of rest time to minimize any fatigue effects from the training.

Borg [11] observed that by providing appropriate verbal anchors and instructions with corresponding muscular sensation, the fatigue ratings can be used to compare inter-individual differences. We adopted Borg's technique and provided detailed instructions on using the Borg CR10 scale with verbal anchoring (see Table 1). The Borg CR10 scale and Table 1 were displayed on screen throughout the experiments.

Exp. 1: Maximum Shoulder Torque Measurements

In this session, we measured the maximum shoulder torque of each subject using two methods: (1) direct measurement using a force sensor and (2) indirect measurement based on an isometric fatiguing exercise.

To directly measure the maximum shoulder torque, we used a dynamometer (MicroFET2, Hoggan Scientific LLC) fixed on a test rig. Figure 4(a) shows an overview of the set-up where the dynamometer (green box) is fixed to a test rig at the same height as the shoulder. Subjects were encouraged to exert maximal upward force with their right arm against the dynamometer. Following the direct maximum torque measurement procedure described by Mathiassen et al. [41], we recorded the best of three trials, each lasting 3 seconds, with a 2 minute interval between repetitions as the maximum force F_{max} . We calculated the maximum shoulder torque as:

$$T_{max,D} = (m_{UA} + m_{LA} + m_H) * G * C + F_{max} * (L_{UA} + L_{LA}),$$

where $T_{max,D}$ is the maximum shoulder torque obtained from the direct force measurement; L_{UA} and L_{LA} are, respectively, the upper and lower arm length; m_{UA} and m_{LA} are, respectively, the mass of the upper and lower arm; m_H

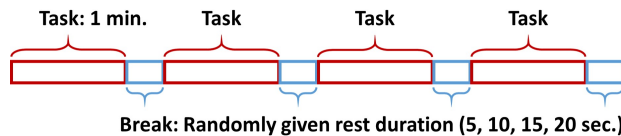


Figure 5. Experimental protocol of pointing task. An experiment block consists of 1 minute of pointing task and either one of 5/10/15/20 seconds of rest, and each pointing task conditions consists of four blocks.

is the mass of the hand; G is 9.81 m/s^2 ; and C is the distance between the shoulder joint and the center of mass of the arm.

In the isometric exercise, we followed the procedure of indirectly estimating maximum shoulder torque described in the previous section. Subjects were asked to hold a weight (2.27 kg for males and 1.36 kg for females, to account for expected gender-based strength differences) in their dominant hand, flex the shoulder till the arm was horizontal and maintain this posture till failure. Subjects saw feedback of their pointing position on screen. We measured the endurance time as the duration between the start of the trial and when subjects indicated volitional failure to maintain the arm posture, or when the hand position was below the target for over 2 seconds. The average torque T_{avg} exerted by their shoulder was computed using inverse dynamics (see the Appendix). Then we obtain the indirectly measured maximum shoulder torque $T_{max,I}$ using Equation 1. During this trial, subjects were required to report their current Borg rating every 20 seconds. These ratings were not utilized for further analysis, but the exercise provided the subjects with Borg rating experience before the next experimental task.

Exp. 2: Mid-Air Pointing Experiments

The main goal of this experiment is to collect subjective fatigue ratings and corresponding biomechanics measures. Then, we aim to generate a cumulative fatigue model that accurately describes the relationship between the two variables. To study this relationship in a principled way, we first need to identify what factors in the interaction design decisively affect cumulative fatigue.

In prior work [30, 31] the authors observed that users get more rapidly fatigued when their arm moves higher from its resting position and extended further away from the body. We chose to investigate the height of the interaction zone as an experimental condition with a fixed range of arm extension, specifically a posture at approximately half of maximum arm stretch. Rest periods are also a decisive factor contributing to cumulative fatigue: the less rest we take, the more we get fatigued. We chose to control these two variables in our study design directly (2×2 experiment conditions). Then we further investigate the applicability of the model for estimating subjective fatigue in different interaction conditions. Figure 4(e) shows the targets arranged on a 2D plane at two different locations (see Figure 4(c),(d)), one at shoulder level and the other midway between the shoulder and waist.

Pilot tests revealed that if subjects perform mid-air tasks in all four conditions (i.e., at least four mid-air pointing tasks in a series), subjects had a higher chance to experience learning and pre-fatigue effects. Thus, we designed our experi-

ment with rest period conditions as a between-subject factor and interaction space conditions as a within-subject one. Each group had 12 participants (2 females), and subjects were given two randomly ordered interaction space conditions. Each group experienced a different order of rest periods – [20s-5s-15s-10s] for group 1 and [5s-10s-20s-15s] for group 2 – during approximately 5 minutes of pointing tasks.

Procedure

We first introduced the entire experimental procedure to the subjects, and surveyed their current physical state and mood along with background information. Then an experimenter measured the body segment parameters of subjects including total weight, height, upper arm length (shoulder to elbow), lower arm length (elbow to wrist), and hand length (wrist to the third finger tip). Based on these measurements, we computed the inertial properties of each upper limb segment based on De Leva's estimation method [16], and used those for the inverse dynamics computations (see Appendix for details).

Next, subjects started the training session for the mid-air pointing tasks. They performed all interaction space conditions in the main experiments with the guidance that they should “hit as many targets as possible while keeping good pointing accuracy”. This is a typical instruction given for Fitts' Law studies, and used to spur the subject's motivation and effort in performing the tasks. We used the same instruction in the second sub-experiment (Exp. 2) to induce and maintain non-trivial arm fatigue during the tasks. After the training session, subjects had at least 5 minutes of rest.

Subsequently, subjects performed the direct maximum shoulder torque measurement. There were three blocks in this measurement consisting of 3 seconds of maximal force exertion and 2 minutes of rest. After the direct force measurement, we provided at least 15 minutes of mandatory rest to ensure no after-task interference, as done in traditional procedures [41]. Then subjects performed the isometric load exercise till they experienced task failure. This was followed by another 30 minutes of mandatory rest.

After this break, we started the mid-air pointing section (Exp. 2). There were two experimental conditions in this session. Each sub-task consists of four task blocks, and each block includes 1 minute of a pointing task and either one of 5, 10, 15, and 20 seconds of rest period (see Fig 5). Subjects took 30 minutes of mandatory rest in between sub-tasks to ensure that there is no interference from prior sub-task. Subjects provided Borg ratings every 20 second during the pointing tasks, and at the start and end of the rest period.

RESULTS

Here we first compare the indirect maximal shoulder torque estimates with the sensor-based direct measurements. Second, we show the effect of rest periods on the Borg ratings. Lastly, we investigate the performance of the TCM model in estimating the subjective fatigue level during mid-air interaction. Table 2 shows the all experiment conditions presented in the mid-air pointing experiments (Exp. 2). In the rest of this section, we denote group 1 and 2 as G1 and G2, and the

	Group 1	Group 2
Interaction zones	G1-H:shoulder level G1-L:waist level	G2-H:shoulder level G2-L:waist level
Rest order	[20s, 5s, 15s, 10s]	[5s, 10s, 20s, 15s]

Table 2. Experimental conditions in the mid-air pointing tasks (Exp. 2).

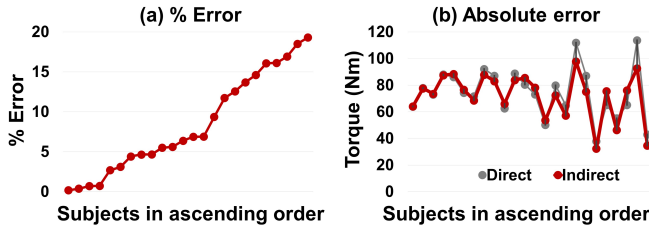


Figure 6. Results of indirect maximum shoulder torque measurement using an isometric fatiguing exercise, sorted by %error across subjects (x-axis). The results are compared with sensor-based direct measures.

high and low interaction zones as H and L. Then, we further analyze the factors on the estimation performance.

Validating Indirect Maximum Muscular Strength Measure

In the isometric load exercise (Exp. 1), we recorded the endurance time and the averaged shoulder torque (T_{avg}). By using these measures in Equation 1, we computed the indirect maximum shoulder torque ($T_{max,I}$). We also recorded the maximum shoulder force exerted toward a dynamometer and computed the direct maximum shoulder torque ($T_{max,D}$) from the force measures. We compared these two measures by evaluating $\%error = \frac{|T_{max,I} - T_{max,D}|}{T_{max,D}} * 100\%$ and $absolute\ error = |T_{max,I} - T_{max,D}|$. Figure 6 shows the errors between $T_{max,I}$ and $T_{max,D}$. Results show only 6.1 Nm (SD=5.0) absolute error and 8.4% (SD=6.21%) relative error. A paired-samples t-Test showed no significant difference between $T_{max,I}$ and $T_{max,D}$ ($t(23)=1.576$, $p=0.129$).

Borg CR10 Scale Ratings across Subjects

Figure 7 shows the averaged Borg ratings for each experimental conditions. In Figure 7(a), G1-H shows a slow development of fatigue relative to G2-H. This is expected, as G1-H has 20 seconds rest in the first task block while G2-H has only 5 seconds rest. Thus, G2-H is a more fatiguing task condition compared to G1-H. A similar trend is visible in Figure 7(b), where G1-L and G2-L are compared. Both rest period conditions ultimately lead to arm muscle fatigue as the average Borg ratings reach their maximum towards the end of the last task block. The introduction of shorter rest periods at the beginning of the tasks seems to affect the overall fatigue level as the highest averaged Borg rating was 6.54 and 3.79 in G1-H and G1-L respectively, while G2-H and G2-L were 6.13 and 3.50. Figure 7 also shows the range of Borg ratings across all subjects. There are large variations across subjects in Borg ratings (average variation range=3.2, min=1, max=6).

Model Performance in Estimating Perceived Exertion

To evaluate the performance of the model in estimating subjective fatigue, we performed multiple cross-validations on interaction conditions and subjects. We also drew a comparison with an existing fatigue metric (CE [30]).

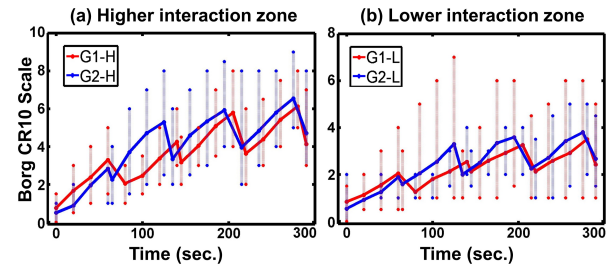


Figure 7. Borg CR10 ratings by 24 subjects during the mid-air pointing tasks. (a) shows the ratings in the high interaction zone in G1-H and G2-H. (b) shows the ratings in the low interaction zone in G1-L and G2-L.

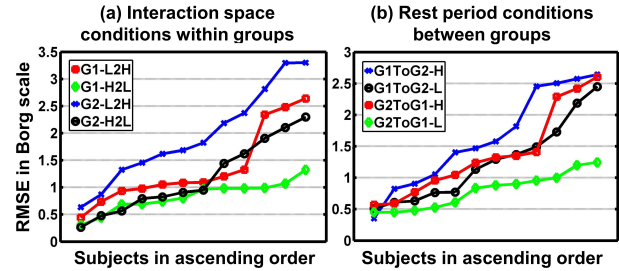


Figure 8. Results of cross-validation across (a): interaction space conditions within groups and (b): rest period conditions between groups. The subjects are sorted by the ascending order of RMSE.

Cross-Validating Interaction Space Conditions

To investigate the applicability of the cumulative fatigue model to different interaction space conditions—higher or lower interaction space—we designed a cross validation test where the model is optimized only with either one of interaction zone conditions. Then we tested it on the data set of the other condition. The cross-validations are performed in each group separately to see the performance in isolation without the potentially confounding effects of rest periods. In total, there were four cross-validations in this experiment, G1-H $\leftrightarrow$ G1-L (2) and G2-H $\leftrightarrow$ G2-L (2). Figure 8(a) depicts the RMSE of each subject data in ascending order. The averaged RMSEs of each cross validation are: G1-L2H = 1.36 (SD=0.72), G1-H2L = 0.83 (SD=0.28), G2-L2H = 1.95 (SD=0.87), and G1-H2L = 1.18 (SD=0.68) on the Borg scale. The overall RMSE is 1.33 (SD=0.77).

Between Subject Groups in Different Conditions

To demonstrate the applicability of the model to different conditions, such as the order of rest period and subjects, we designed a cross validation where the model is optimized using either one of two group's data. Then the model is tested with the other group's data. In this validation, we can evaluate how well the model can be used in different situations. There were total four cross-validations: G2 $\rightarrow$ G1-H/L (2) and G1 $\rightarrow$ G2-H/L (2). Figure 8(b) shows the between-group cross-validation results. The averaged RMSEs of each cross validation are: 'G2ToG1-H' = 1.38 (SD=0.70), 'G2ToG1-L' = 0.79 (SD=0.29), 'G1ToG2-H' = 1.63 (SD=0.78), and 'G2ToG1-L' = 1.24 (SD=0.64) on the Borg scale. The overall RMSE is 1.26 (SD=0.35), which is slightly lower than the cross-validation between interaction space conditions (1.33). To further investigate the statistical difference between the

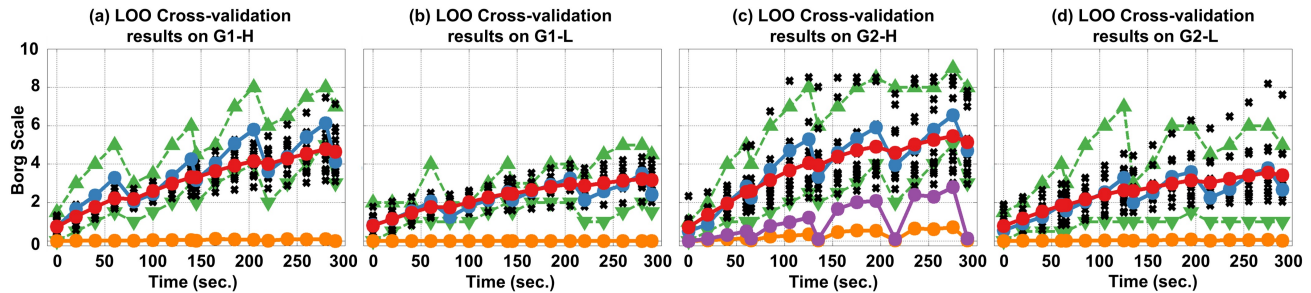


Figure 9. Results of the leave-one-out cross-validation of the model across all subject data. Green-upward/downward triangles: upper/lower bound of ground truth, blue circles: averaged ground truth, black crosses: TCM estimates, red circles: averaged TCM estimates, orange circles: averaged CE [30] estimates, purple circles: averaged CE estimates on the subject data with average %MVC above 15% (only observed in G2-H condition).

cross validation results between groups, we performed a paired t-test. Results showed no significant difference in the averaged RMSE ($t(47)=-0.713$, $p=0.479$). Thus, we conclude that the between group conditions did not affect the estimation performance of the model.

Effect of Interaction Conditions on Fatigue Estimation

To investigate the effect of task conditions—interaction zones and rest period order—on the estimation performance measured in RMSE, we performed a mixed factors ANOVA test with *rest period order* as between-subject factor and *interactions zones* as within-subject factor. We optimized the model using all data collected from the experiments (24 subjects $\times$ 2 interaction tasks), and obtained the model parameters as $F = 0.0146$ and $R = 0.0022$. The ANOVA results showed that there is no significant interaction effect between rest order and interaction zone ($F_{1,22} = 0.026$, $p = 0.874$); no significant difference between rest order conditions ($F_{1,22} = 2.361$, $p = 0.390$); but also that the interaction zone conditions are significantly different ($F_{1,22} = 26.812$, $p < 0.001$). Overall, the upper interaction zone (RMSE-M=1.52, SD=0.76) showed higher fatigue estimation errors compared to the lower condition (RMSE-M=0.98, SD=0.54).

The parameter optimization discussed above is ecologically less sound since it was performed using all subject data. To illustrate the robustness of our results, we decided to perform a leave-one-out (LOO) cross-validation for the fatigue models generated for each subject. For example, the first subject's data is excluded from optimizing the model, so that we can also test the model toward unseen subject's data. We assume that the RMSE measured using each model is independent from each other (i.e., only one data is eliminated in the optimization, and it does not affect the model parameters). We performed the same mixed factors ANOVA test, but with the models generated from the LOO cross-validations. With this, we obtained similar results in that there is no significant interaction effect between rest order and interaction zone ($F_{1,22} = 0.109$, $p = 0.745$); no significant effect of the two rest order conditions ($F_{1,22} = 3.589$, $p = 0.071$); and a significant effect of interaction zone conditions ($F_{1,22} = 17.350$, $p < 0.001$). Similar to the previous ANOVA test, the upper interaction zone (mean RMSE=1.50, SD=0.74) showed higher fatigue estimation errors compared to the lower condition (mean RMSE=1.03, SD=0.56). Based on these two congruent ANOVA test results, we conclude that the estimation

performance of the TCM model is less affected by changes of rest duration compared to interaction space conditions.

Comparison with Consumed Endurance Metric

To compare the estimation performance of our model with an existing fatigue metric, *Consumed Endurance* (CE [30]), we computed CE on all subject data used in the LOO cross-validation test. For CE, we used each subject's maximum shoulder torque and the torque profile measured during the pointing tasks. The averaged %MVC ($= T/T_{max} \times 100\%$) for each subject observed during the pointing tasks was 11.79% (Range=[8.11,17.43]). We converted CE estimates to the Borg scale with $CE \times 0.1 = Borg$, as identified by Hincapié-Ramos et al. [30]. The *total interaction time* for CE included both active and rest periods. As CE only quantifies instantaneous fatigue, we used averaged CE values of each period in between Borg ratings to estimate cumulative fatigue.

Figure 9 shows an overview of the results for the TCM model (ours) and CE across all subject data. The averaged TCM estimates (red circles) mostly follow the trend of averaged Borg ratings (blue circles). Also, individual Borg estimates (black crosses) are well within the min-max bounds (green triangles) of the Borg ratings. As CE automatically yields a fatigue level of zero for behaviors that induce exertion below 15%MVC, CE fails to properly quantify the fatigue level in all conditions. As our tasks mostly involved exertion below this threshold (avg.=11.79%MVC), this causes CE to underestimate fatigue. This fundamental limitation of CE yields larger overall errors compared to TCM in all task conditions (TCM: overall-RMSE=1.46, Range=[0.83,1.90], CE: overall-RMSE=3.45, Range=[2.42,4.25]).

We also compared our model with CE specifically for the data with averaged %MVC higher than 15% (N=4, avg. %MVC-Mean=15.93%, Range=[15.01,17.43]). Our TCM model (overall-RMSE=0.93, Range=[0.67,1.19]) still showed lower error than CE (overall-RMSE=2.96, Range=[2.12,3.60]). As shown in Figure 9(c) (purple circles), CE tends to almost monotonously increase while underestimating fatigue levels. Also, CE estimates drop suddenly close to zero during rest periods, failing to capture the continuity of fatigue.

In contrast to CE, our TCM model does not limit the range of exertion when estimating cumulative fatigue. Also, our model considers both rest and active periods, and correctly predicts cumulative fatigue for intermittent task periods.

Inter-individual Factors on Model Performance

We examined the potential effect of (1) subject's background (e.g., age, overall tiredness, experience with mid-air interfaces), (2) physical measurements (weight, height, limb length), and (3) variables collected during the maximum torque estimation (errors in the estimation, endurance time, maximum torque) on the errors in estimating subjective fatigue. We performed a Pearson's product-moment test with these individual variables, and none of them showed a significant correlation with the estimation errors. We further examined the relationship between Borg ratings and shoulder joint torque. These two parameters are directly used in both optimizing and evaluating the fatigue model, so it is worth investigating their effects on the estimation performance. In particular, we examined the ratio ($r_{B/T}$) between Borg ratings and shoulder joint torque during the mid-air pointing tasks: $r_{B/T} = \frac{Borg_{avg}}{T_{avg}}$, where $Borg_{avg}$ is the average of Borg ratings and T_{avg} is the average of shoulder torques collected from a subject performing a mid-air pointing task. We performed a linear regression analysis between $r_{B/T}$ and RMSE measured in the leave-one-out cross validation with all data. Results showed that there was a strong correlation between the estimation error and $r_{B/T}$ ($F_{1,46} = 109.148$, $p < 0.001$) with an R^2 of 0.704. A multiple linear regression was further calculated to predict the RMSE based on both factors $Borg_{avg}$ and T_{avg} . Results showed that the two factors explained 70.4% of the variance ($R^2 = 0.704$, $F_{2,45} = 53.401$, $p < 0.001$). However, only $Borg_{avg}$ was a significant predictor ($\beta = 0.687$, $p < 0.001$). This can be explained by (1) a significant correlation between subject's internal variability of Borg ratings and estimation errors and (2) limitations of the linear function ($\phi(\cdot)$) mapping Borg ratings to the estimate fatigue level M_F .

DISCUSSION

Our results validated the value of our approach for estimating subjective fatigue only using a camera-based skeleton tracking system instead of expensive and hard-to-access methods that are impractical for HCI.

Simple and Effective Shoulder Torque Measures

A comparison between our indirect maximum shoulder torque measurement through a simple mid-air pointing task showed no significant difference compared to traditional methods. To our knowledge, this is the first method to estimate an individual's maximum shoulder torque with only vision-based skeleton tracking sensors. We envision that our simple but effective method can be used in various HCI scenarios. For example, personalized training programs for Kinect-based at-home therapy applications [57] could be adapted to each individual's strength using our method.

Reliable Performance in Estimating Subjective Fatigue

Our results showed good estimation of subjective fatigue (maximum-mean-RMSE=1.95 (SD=0.87) observed in G2-L2H, minimum-mean-RMSE=0.79 (SD=0.29) observed in G2ToG1-L). Given the variability in subjective evaluations of fatigue, we believe that our model did well in generating such low errors. This approach is a significant stepping-stone

to investigate the relation between subjective and objective fatigue. The computational complexity for implementing our algorithm is low, and it is implementable in real-time (50Hz).

Limitations and Future Work

While our model showed overall good estimation performance (overall-RMSE=1.46 in the LOO cross-validation), we identified some limitations that should be addressed in future work. First, our proposal is based on an assumption that through properly transforming biomechanical information we can estimate the perceived exertion level. However, as subjective fatigue is influenced by both physiological and psychological changes, there are several potential factors that could not be modeled with our two parameters. In the experiments, we observed that some subjects indicated much higher Borg ratings than what one might expect. For example, subject 19 reported about 8 to 10 %MVC while performing the pointing tasks, but indicated the highest Borg ratings—expressed in his native language—among all participants. This suggests that this subject may experience fatigue and rest differently compared to others (i.e., beyond the normal range for F and R parameters). Such inter-individual difference should be further studied to reduce estimation errors.

Second, our current assumption of a linear mapping between the perceived exertion and the maximum joint force should be revisited. Some psychophysical relations between muscular intensity and perceived exertion ratings can be described by power functions [12]. In the future, we will focus on identifying such a mapping function for a focus group of subjects.

We also plan to extend our approach to broader and complex interaction scenarios. For example, we can apply the TCM model to other mid-air interaction tasks (e.g., rotation, docking) by fitting the model parameters, as long as the corresponding data (joint torque, Borg ratings) is available. In general, fatigue is task-dependent [19, 20], and different model parameters might be needed for different tasks.

We are also looking at testing the model on tasks that involve highly-trained movements during interaction with varying rest periods, such as playing gesture-based games. Our model already quantifies the effect of rest on cumulative fatigue well, but it might be possible to improve estimation accuracy for varying rest periods via weighted optimization.

CONCLUSION

We presented a framework for evaluating subjective fatigue using only a vision-based skeleton tracking system. First, we proposed an indirect method to estimate the maximum shoulder torque through a simple mid-air pointing task. Comparison with traditional sensor-based direct measures showed no significant difference between the two methods. Second, we applied the three-compartment muscle model for estimating subjective arm fatigue. To our best knowledge, this is the first attempt to quantify cumulative subjective arm fatigue in a continuous task involving rest periods using only biomechanical information generated by a vision system. Overall, our model showed good estimation performance in multiple cross-validations. Statistical analysis revealed that the model's performance was robust to changes in the duration

Link	θ_i	α_i	m_i	${}^i\vec{S}_i$	${}^i\vec{I}_i$
1	θ_1	$-\frac{\pi}{2}$	0	$\vec{0}$	$0_{3 \times 3}$
2	$\theta_2 - \frac{\pi}{2}$	$\frac{\pi}{2}$	0	$\vec{0}$	$0_{3 \times 3}$
3	$\theta_3 + \frac{\pi}{2}$	$\frac{\pi}{2}$	m_{UA}	$[0, d_{UA}, 0]^T$	$diag(I_{UAx}, I_{UAz}, I_{UAy})$
4	θ_4	0	$m_{LA} + m_H$	$[d_4, 0, 0]^T$	$diag(I_{4z}, I_{4x}, I_{4y})$

Table 3. D-H and body segment parameters for 4-DOF arm model (*: find detailed definition in the paragraph below).

of rest, but relatively less robust to changes of the interaction zone. Moreover, our model outperformed an existing fatigue metric (CE [30]) as ours has no limits on the range of exertion considered and takes rest correctly into account.

ACKNOWLEDGEMENTS

We thank the reviewers for their valuable feedback. This work was partly supported by the National Science Foundation (Award # 1538868 CMMI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF.

Source Code Availability

The fatigue model implementation and the biomechanical upper limb analysis will be released on the project website: <http://www.tinyurl.com/cumulative-arm-fatigue>

APPENDIX: BIOMECHANICAL ARM MODEL

In Figure 3, four cylinders represent 1-DOF revolute joints ($J_0 \sim J_3$), and vectors $\vec{z}_0 \sim \vec{z}_3$ indicate their respective rotation axis. The first three joints ($J_0 \sim J_2$) share the same position at the shoulder joint, and J_3 is attached at the elbow joint. Thus, the arm is a 4-DOF serial chain (shoulder:3, elbow:1). J_i connects i^{th} and $(i+1)^{\text{th}}$ link, and the link 0 is fixed to the ground body. The coordinate frame 0: $(\vec{x}_0, \vec{y}_0, \vec{z}_0)$ is the base, and its origin is at J_0 . The frame i : $(\vec{x}_i, \vec{y}_i, \vec{z}_i)$ is attached to link i , and its origin is at J_i ($i \in [1, 2, 3]$). Lastly, the frame 4: $(\vec{x}_4, \vec{y}_4, \vec{z}_4)$ is fixed at the end of the lower arm (wrist joint).

Kinematic Analysis

We use De Leva's standard approach [16] to estimate the body segment parameters (BSPs) of the upper arm (UA), lower arm (LA), and hand (H), including the mass (m_{UA} , m_{LA} , m_H), the center of mass (COM) distance (d_{UA} , d_{LA} , d_H), and the moment of inertia matrix ($\in \mathbb{R}^{3 \times 3}$) of the arm segments ($\vec{I}_{UA} = diag(I_{UAx}, I_{UAy}, I_{UAz})$, $\vec{I}_{LA} = diag(I_{LAx}, I_{LAy}, I_{LAz})$, $\vec{I}_H = diag(I_{Hx}, I_{Hy}, I_{Hz})$) based on each individual's upper-arm (L_{UA}), forearm (L_{LA}), and hand (L_H) lengths, and total mass of body (m_{Tot}). We compute the COM distances by $d_{UA} = L_{UA}(100 - \bar{L}_{UA})$, $d_{LA} = -L_{LA}(100 - \bar{L}_{LA})$, $d_H = L_H * \bar{L}_H$; and body part masses by $m_{UA} = m_{Tot} * \bar{m}_{UA}$, $m_{LA} = m_{Tot} * \bar{m}_{LA}$, $m_H = m_{Tot} * \bar{m}_H$, where $\bar{L}_i$ s and $\bar{m}_i$ s are, respectively, gender dependent %length and %mass parameters [16].

The orientation of each joint axis is specified using the Denavit-Hartenberg (D-H) notation [18]. Table 3 defines the D-H parameters for the 4-DOF arm model, where θ_i is the rotation angle of i^{th} link about $\vec{z}_{i-1}$ axis, and α_i is the angle

between $\vec{z}_{i-1}$ and $\vec{z}_i$ about $\vec{x}_{i-1}$ axis. Based on the D-H parameters, the rotational matrices among the coordinate frames are constructed as:

$${}^{i-1}R_i = \begin{bmatrix} \cos(\theta_i) & -\cos(\alpha_i)\sin(\theta_i) & \sin(\alpha_i)\sin(\theta_i) \\ \sin(\theta_i) & \cos(\alpha_i)\cos(\theta_i) & -\sin(\alpha_i)\cos(\theta_i) \\ 0 & \sin(\alpha_i) & \cos(\alpha_i) \end{bmatrix}$$

where ${}^{i-1}R_i \in \mathbb{R}^{3 \times 3}$ is a rotation matrix transforming any vector in the i^{th} frame into the $(i-1)^{\text{th}}$ frame ¹.

Newton-Euler Inverse Dynamics

To compute joint torques from a given arm motions represented by the biomechanics model, we use a recursive Newton-Euler dynamics formulation (RNE) [15]. It first computes velocities and accelerations of each link from the 1st to 4th link, and then computes joint forces and torques in reverse order. Thus, the effects of forces and torques are propagated from the last (distal) to the first (proximal) link: Outward recursion ($1 \rightarrow 4$):

$$\begin{aligned} {}^i\vec{\omega}_i &= {}^iR_{i-1} [{}^{i-1}\vec{\omega}_{i-1} + \hat{z}\dot{\theta}_i], \\ {}^i\dot{\vec{\omega}}_i &= {}^iR_{i-1} [{}^{i-1}\dot{\vec{\omega}}_{i-1} + \hat{z}_0\ddot{\theta}_i + {}^{i-1}\vec{\omega}_{i-1} \times \hat{z}_0\dot{\theta}_i], \\ {}^i\vec{v}_i &= {}^i\vec{\omega}_i \times {}^i\vec{P}_i^* + {}^i\vec{\omega}_i \times [{}^i\vec{\omega}_i \times {}^i\vec{P}_i^*] + {}^iR_{i-1} {}^{i-1}\vec{v}_{i-1}, \\ {}^i\vec{a}_i &= {}^i\dot{\vec{\omega}}_i \times {}^i\vec{S}_i + {}^i\vec{\omega}_i \times [{}^i\vec{\omega}_i \times {}^i\vec{S}_i] + {}^i\vec{v}_i \end{aligned}$$

Inward recursion ($4 \rightarrow 1$):

$$\begin{aligned} {}^i\vec{f}_i &= {}^iR_{i+1} {}^{i+1}\vec{f}_{i+1} + m_i {}^i\vec{a}_i, \\ {}^i\vec{n}_i &= {}^iR_{i+1} [{}^{i+1}\vec{n}_{i+1} + {}^{i+1}R_i \vec{P}_i^* \times {}^{i+1}\vec{f}_{i+1}] \\ &\quad + [{}^i\vec{P}_i^* \times {}^i\vec{S}_i] \times (m_i {}^i\vec{a}_i) + {}^i\vec{I}_i {}^i\dot{\vec{\omega}}_i + {}^i\vec{\omega}_i \times ({}^i\vec{I}_i {}^i\vec{\omega}_i) \end{aligned}$$

where $\hat{z} = [0, 0, 1]^T$; ${}^i\vec{\omega}_i$ and ${}^i\dot{\vec{\omega}}_i$ are, respectively, the angular velocity and acceleration vectors of the i^{th} link ²; ${}^i\vec{v}_i$ and ${}^i\vec{a}_i$ are, respectively, the linear acceleration vectors of the origin of the i^{th} frame and COM for the i^{th} link; ${}^i\vec{f}_i$ and ${}^i\vec{n}_i$ are, respectively, the force and moment exerted on the i^{th} link by the $(i-1)^{\text{th}}$ link; ${}^i\vec{P}_i^*$ is a vector linking the origin of the $(i-1)^{\text{th}}$ and i^{th} frame; ${}^i\vec{S}_i$ is a vector linking the origin of the i^{th} frame and COM of the i^{th} link; m_i is the mass of the i^{th} link; and ${}^i\vec{I}_i \in \mathbb{R}^{3 \times 3}$ is a diagonal inertia matrix. The last three variables are constant BSPs and defined in Table 3. We calculate BSPs of link 4 by combining forearm and hand as one rigid body: $m_4 = m_{LA} + m_H$, $d_4 = \frac{d_{LA} * m_{LA} + d_H * m_H}{m_{LA} + m_H}$, $I_{4x} = I_{LAx} + I_{Hx}$, $I_{4y} = I_{LAy} + m_{LA}(d_4 - d_{LA})^2 + I_{Hy} + m_H(d_4 - d_H)^2$, $I_{4z} = I_{LAz} + m_{LA}(d_4 - d_{LA})^2 + I_{Hz} + m_H(d_4 - d_H)^2$. We set the initial values as ${}^0\vec{\omega}_0 = {}^0\dot{\vec{\omega}}_0 = {}^5\vec{f}_5 = {}^5\vec{n}_5 = \vec{0}$ and ${}^0\vec{v}_0 = [-9.81, 0, 0]^T$ m/s². Finally, we compute the shoulder torque as:

$$T = \sqrt{\tau_0 + \tau_1 + \tau_2},$$

where $\tau_i = [({}^{i+1}\vec{n}_{i+1})^T ({}^{i+1}R_i \hat{z})]^2$ and $\hat{z} = [0, 0, 1]^T$.

¹ $[{}^{i-1}R_i]^{-1} = [{}^{i-1}R_i]^T = {}^iR_{i-1}$

²A leading superscript F of a vector ${}^F\vec{x} \in \mathbb{R}^3$ indicates that the vector is referenced to the coordinate frame F .

REFERENCES

1. Biodex dynamometer. <http://www.biodex.com/physical-medicine/products/dynamometers>.
2. Gorilla arm. <http://www.computer-dictionary-online.org/?q=gorilla%20arm>.
3. Amann, M., Romer, L. M., Pegelow, D. F., Jacques, A. J., Hess, C. J., and Dempsey, J. A. Effects of arterial oxygen content on peripheral locomotor muscle fatigue. *Journal of Applied Physiology* 101, 1 (2006), 119–127.
4. Åstrand, P.-O. *Textbook of work physiology: physiological bases of exercise*. Human Kinetics, 2003.
5. Audet, C., and Dennis Jr, J. E. Analysis of generalized pattern searches. *SIAM Journal on Optimization* 13, 3 (2002), 889–903.
6. Bachynskyi, M., Oulasvirta, A., Palmas, G., and Weinkauff, T. 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 (2014), 3215–3224.
7. Bachynskyi, M., Palmas, G., Oulasvirta, A., Steimle, J., and Weinkauff, T. 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 (2015), 1817–1826.
8. Bachynskyi, M., Palmas, G., Oulasvirta, A., and Weinkauff, T. Informing the design of novel input methods with muscle coactivation clustering. *ACM Trans. Comput.-Hum. Interact.* 21, 6 (Jan. 2015), 30:1–30:25.
9. Bijur, P. E., Silver, W., and Gallagher, E. J. Reliability of the visual analog scale for measurement of acute pain. *Academic emergency medicine* 8, 12 (2001), 1153–1157.
10. Borg, G. Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian journal of work, environment & health* (1990), 55–58.
11. Borg, G. *Borg's perceived exertion and pain scales*. Human kinetics, 1998.
12. Borg, G. A. Psychophysical bases of perceived exertion. *Med sci sports exerc* 14, 5 (1982), 377–381.
13. Carifio, J., and Perla, R. J. Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences* 3, 3 (2007), 106–116.
14. Cifrek, M., Medved, V., Tonković, S., and Ostojić, S. Surface emg based muscle fatigue evaluation in biomechanics. *Clinical Biomechanics* 24, 4 (2009), 327–340.
15. Craig, J. J. *Introduction to robotics: mechanics and control*, vol. 3. Pearson Prentice Hall Upper Saddle River, 2005.
16. De Leva, P. Adjustments to zatsiorsky-seluyanov's segment inertia parameters. *Journal of biomechanics* 29, 9 (1996), 1223–1230.
17. Dederer, Å., Németh, G., and Harms-Ringdahl, K. Correlation between electromyographic spectral changes and subjective assessment of lumbar muscle fatigue in subjects without pain from the lower back. *Clinical Biomechanics* 14, 2 (1999), 103–111.
18. Denavit, J., and Hartenberg, R. S. A kinematic notation for lower-pair mechanisms based on matrices. *Trans. ASME, J. Appl. Mech.* 22, 2 (1965), 215 – 221.
19. Enoka, R. M., and Duchateau, J. Muscle fatigue: what, why and how it influences muscle function. *The Journal of physiology* 586, 1 (2008), 11–23.
20. Enoka, R. M., and Stuart, D. G. Neurobiology of muscle fatigue. *Journal of applied physiology* 72, 5 (1992), 1631–1648.
21. Eston, R. Use of ratings of perceived exertion in sports. *International journal of sports physiology and performance* 7, 2 (2012), 175–182.
22. Freivalds, A. *Biomechanics of the upper limbs: mechanics, modeling and musculoskeletal injuries*. CRC press, 2011.
23. Frey-Law, L. A., and Avin, K. G. Endurance time is joint-specific: a modelling and meta-analysis investigation. *Ergonomics* 53, 1 (2010), 109–129.
24. Frey-Law, L. A., Looft, J. M., and Heitsman, J. A three-compartment muscle fatigue model accurately predicts joint-specific maximum endurance times for sustained isometric tasks. *Journal of biomechanics* 45, 10 (2012), 1803–1808.
25. Haga, S., Shinoda, H., and Kokubun, M. Effects of task difficulty and time-on-task on mental workload. *Japanese Psychological Research* 44, 3 (2002), 134–143.
26. Harrison, C., Ramamurthy, S., and Hudson, S. E. On-body interaction: armed and dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, ACM (2012), 69–76.
27. Hart, S. G., and Staveland, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183.
28. Hassenzahl, M., and Tractinsky, N. User experience-a research agenda. *Behaviour & information technology* 25, 2 (2006), 91–97.
29. Hayes, K., Walton, J. R., Szomor, Z. L., and Murrell, G. A. Reliability of 3 methods for assessing shoulder strength. *Journal of Shoulder and Elbow surgery* 11, 1 (2002), 33–39.

30. Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., and Irani, P. 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 (2014), 1063–1072.
31. Hofmann, M., Brger, R., Frost, N., Karremann, J., Keller-Bacher, J., Kraft, S., Bruder, G., and Steinicke, F. Comparing 3d interaction performance in comfortable and uncomfortable regions. In *Proceedings of the GI-Workshop VR/AR* (2013), 3–14.
32. Hornbæk, K., and Law, E. L.-C. Meta-analysis of correlations among usability measures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2007), 617–626.
33. Höysniemi, J., Hämäläinen, P., and Turkki, L. Wizard of oz prototyping of computer vision based action games for children. In *Proceedings of the 2004 conference on Interaction design and children: building a community*, ACM (2004), 27–34.
34. Hunter, S. K., Duchateau, J., and Enoka, R. M. Muscle fatigue and the mechanisms of task failure. *Exercise and sport sciences reviews* 32, 2 (2004), 44–49.
35. ISO, I. 9241-9 ergonomic requirements for office work with visual display terminals (vdts)-part 9: Requirements for non-keyboard input devices (fdis-final draft international standard), 2000. *International Organization for Standardization*.
36. Jacob, M. G., and Wachs, J. P. Context-based hand gesture recognition for the operating room. *Pattern Recognition Letters* 36 (2014), 196–203.
37. Kaiser, E., Olwal, A., McGee, D., Benko, H., Corradini, A., Li, X., Cohen, P., and Feiner, S. Mutual disambiguation of 3d multimodal interaction in augmented and virtual reality. In *Proceedings of the 5th international conference on Multimodal interfaces*, ACM (2003), 12–19.
38. Liu, J. Z., Brown, R. W., and Yue, G. H. A dynamical model of muscle activation, fatigue, and recovery. *Biophysical journal* 82, 5 (2002), 2344–2359.
39. Lloyd, A., Gandevia, S., and Hales, J. Muscle performance, voluntary activation, twitch properties and perceived effort in normal subjects and patients with the chronic fatigue syndrome. *Brain* 114, 1 (1991), 85–98.
40. Malik, S., McDonald, C., and Roth, G. Hand tracking for interactive pattern-based augmented reality. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, IEEE (2002), 117–126.
41. Mathiassen, S. E., and Åhsberg, E. Prediction of shoulder flexion endurance from personal factors. *International Journal of Industrial Ergonomics* 24, 3 (1999), 315–329.
42. Morishita, S., Yamauchi, S., Fujisawa, C., and Domen, K. Rating of perceived exertion for quantification of the intensity of resistance exercise. *International Journal of Physical Medicine & Rehabilitation* 2013 (2014).
43. Noble, B. J. Clinical applications of perceived exertion. *Medicine and science in sports and exercise* 14, 5 (1981), 406–411.
44. Öberg, T., Sandjö, L., and Kadefors, R. Subjective and objective evaluation of shoulder muscle fatigue. *Ergonomics* 37, 8 (1994), 1323–1333.
45. Raikova, R. A general approach for modelling and mathematical investigation of the human upper limb. *Journal of biomechanics* 25, 8 (1992), 857–867.
46. Rohmert, W. Ermittlung von erholungspausen für statische arbeit des menschen. *European Journal of Applied Physiology and Occupational Physiology* 18, 2 (1960), 123–164.
47. Roy, J.-S., Ma, B., MacDermid, J. C., and Woodhouse, L. J. Shoulder muscle endurance: the development of a standardized and reliable protocol. *BMC Sports Science, Medicine and Rehabilitation* 3, 1 (2011), 1.
48. Segerstrom, S. C., and Nes, L. S. Heart rate variability reflects self-regulatory strength, effort, and fatigue. *Psychological science* 18, 3 (2007), 275–281.
49. Sjøgaard, G., Savard, G., and Juel, C. Muscle blood flow during isometric activity and its relation to muscle fatigue. *European journal of applied physiology and occupational physiology* 57, 3 (1988), 327–335.
50. Sonne, M. W., and Potvin, J. R. A modified version of the three-compartment model to predict fatigue during submaximal tasks with complex force-time histories. *Ergonomics* 59, 1 (2016), 85–98.
51. Spielholz, P. Calibrating borg scale ratings of hand force exertion. *Applied Ergonomics* 37, 5 (2006), 615–618.
52. Teather, R., and Stuerzlinger, W. Pointing at 3D targets in a stereo head-tracked virtual environment. In *Proceedings of the Symposium on 3D User Interfaces*, IEEE (2011), 87–94.
53. Troiano, A., Naddeo, F., Sossio, E., Camarota, G., Merletti, R., and Mesin, L. Assessment of force and fatigue in isometric contractions of the upper trapezius muscle by surface emg signal and perceived exertion scale. *Gait & posture* 28, 2 (2008), 179–186.
54. Vinayak, Murugappan, S., Liu, H., Ramani, K., et al. Shape-it-up: Hand gesture based creative expression of 3d shapes using intelligent generalized cylinders. *Computer-Aided Design* 45, 2 (2013), 277–287.
55. Westerblad, H., Allen, D. G., and Lännergren, J. Muscle fatigue: lactic acid or inorganic phosphate the major cause? *Physiology* 17, 1 (2002), 17–21.
56. Xia, T., and Law, L. A. F. A theoretical approach for modeling peripheral muscle fatigue and recovery. *Journal of biomechanics* 41, 14 (2008), 3046–3052.
57. Zhao, W., Feng, H., Lun, R., Espy, D. D., and Reinthal, M. A. A kinect-based rehabilitation exercise monitoring and guidance system. In *Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on*, IEEE (2014), 762–765.