Weight-Induced Consumed Endurance (WICE): A Model to Quantify Shoulder Fatigue with Weighted Objects

Tinghui Li
School of Computer Science
University of Sydney
Sydney, Australia
tinghui.li@sydney.edu.au

Eduardo Velloso
School of Computer Science
University of Sydney
Sydney, Australia
eduardo.velloso@sydney.edu.au

Anusha Withana
School of Computer Science
University of Sydney
Sydney, Australia
anusha.withana@sydney.edu.au

Zhanna Sarsenbayeva
School of Computer Science
University of Sydney
Sydney, Australia
zhanna.sarsenbayeva@sydney.edu.au

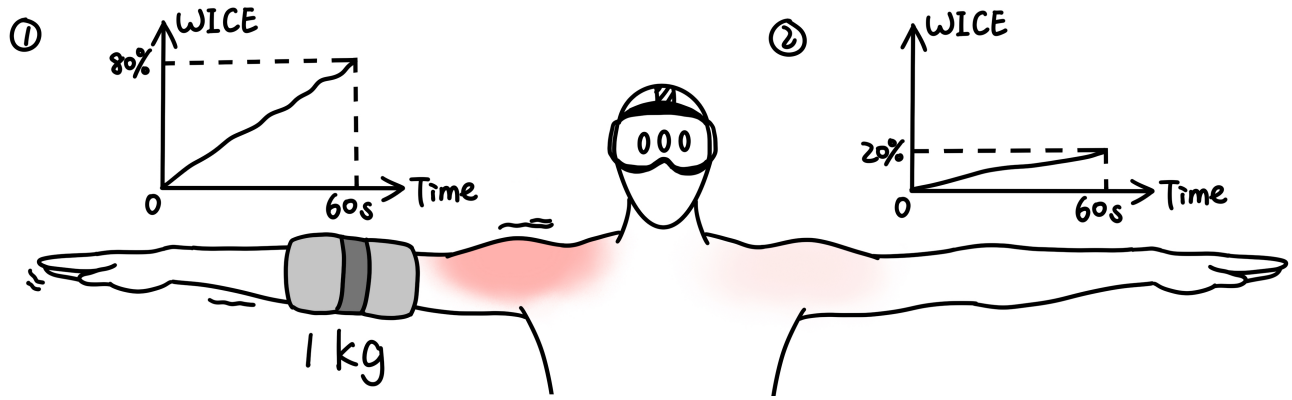


Figure 1: The consumed endurance at 60 seconds. (1) attaching 1 kg weight boosts arm exertion to 80%; (2) bare-hand only consumed 20%. A more intense red coloration on the shoulder indicates a higher level of fatigue experienced by the user.

Abstract

Fatigue is a major challenge in mid-air interactions, often resulting in a sensation of heaviness—particularly when users carry weighted objects on their arms. Existing models for characterising shoulder fatigue were primarily developed for bare-hand scenarios, limiting their applicability in situations involving encumbrance. In this paper, we introduce Weight-Induced Consumed Endurance (WICE), a novel model that accurately estimates shoulder fatigue when additional weight is attached at various locations on the arm. WICE enhances the calculation of instantaneous shoulder torque by incorporating information about the attached weight, integrates individual arm mass for more personalised fatigue estimation, and uses a Bayesian framework to simulate the distribution of shoulder fatigue. Our evaluation shows that WICE strongly correlates with both experimentally measured endurance time and subjective Borg CR10 ratings, demonstrating its reliability as an objective fatigue metric in both encumbered and no-weight conditions. We further demonstrate how WICE can be applied to examine the effects of controller and haptic devices on user fatigue. WICE provides a foundation for developing fatigue-aware systems that can sense and adapt encumbrance, allowing for more tailored ergonomic MR interactions.

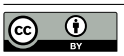
CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; **Virtual reality**; **HCI theory, concepts and models**; **Systems and tools for interaction design**.

Keywords

Endurance, Consumed Endurance, Fatigue, Ergonomics, Encumbrance, Mid-Air Interactions, Mixed Reality

ACM Reference Format:
Tinghui Li, Eduardo Velloso, Anusha Withana, and Zhanna Sarsenbayeva. 2025. Weight-Induced Consumed Endurance (WICE): A Model to Quantify Shoulder Fatigue with Weighted Objects. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28–October 01, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3746059.3747745



UIST '25, Busan, Republic of Korea

ACM ISBN 979-8-4007-2037-6/25/09
https://doi.org/10.1145/3746059.3747745

1 Introduction

Predicting and effectively managing user fatigue during long interactions in immersive settings like mixed and virtual reality is crucial. Manufacturers suggest taking breaks every 30 minutes to prevent fatigue[1], but hardly anyone can continuously interact in mid-air for that long [19, 29, 32]. Previous works have proposed quantitative models for estimating shoulder fatigue in mid-air interaction [34, 36, 42, 43, 50, 71], but these works overlook a key issue—*encumbrance*. This includes the physical load from wearable devices, controllers and other equipment used in mixed reality (MR). Accounting for encumbrance is crucial because MR interactions involve controllers and wearables with weights of their own that also affect participants' ability to perform mid-air gestures. Moreover, encumbrance is also prevalent in the professional and industrial settings in which MR has been used, such as when wearing safety gear[2], protective gloves[3], or carrying firefighting equipment[4]. Finally, if MR technology aims to achieve widespread adoption in daily life, there will be scenarios where users engage with MR devices while carrying objects, such as during shopping or commuting.

To address this gap, in this paper, we propose a novel model for predicting both real-time and cumulative shoulder fatigue during mid-air interactions called **Weight-Induced Consumed Endurance (WICE)**. This is the first model to accurately characterise shoulder fatigue while accounting for the effect of carried objects. WICE improves instantaneous shoulder torque calculations by incorporating carried object information while remaining effective even in no-weight conditions. It also integrates individual arm mass information for more personalised shoulder fatigue calculation. WICE was implemented using a Bayesian framework to simulate the distribution of shoulder fatigue, allowing for probabilistic rather than single-point estimates. We validated that WICE correlated strongly with Borg CR10 and accurately predicted empirical endurance times, achieving the best predictive performance among existing models [34, 42, 43]. We further demonstrate how WICE can be applied to examine the effects of the controller on interaction fatigue, and provide a foundation for sensing and adapting to the effects of encumbrance for MR interactions.

Overall, our study contributes to modelling shoulder fatigue in mid-air interactions by considering the weight attached at various arm locations. The contributions of this paper are as follows:

- We develop a Weight-Induced Consumed Endurance (WICE) model to accurately characterise shoulder fatigue with carried objects by enhancing instantaneous shoulder torque calculations and integrating individual mass information to replace average mass for more personalised calculation;
- We incorporate a Bayesian framework to simulate the distribution of shoulder fatigue, allowing for probabilistic rather than single-point estimates;
- We demonstrate how WICE is useful for informing future interaction designs, particularly in scenarios with carried objects.

[1]https://www.meta.com/au/legal/quest/health-and-safety-warnings/quest-3/ [Accessed: 2025-03-31]
[2]https://health.ucdavis.edu/news/headlines/the-future-of-surgery-using-augmented-reality-goggles-in-the-operating-room/2024/08 [Accessed: 2025-03-31]
[3]https://deadeyevr.com/products/ultimate-boxing-gloves-quest [Accessed: 2025-03-31]
[4]https://youtu.be/UlNQiLfAYhY?si=dtFWjzE_qXuFLA1J [Accessed: 2025-03-31]

2 Related Work

2.1 Subjective Fatigue Measures

Subjective instruments are often used to measure fatigue levels, with the most popular approach being the Borg Category-Ratio scale (**Borg CR10**) categorising physical exertion using discrete levels ranging from 0 (no exertion) to 10 (maximum exertion) [12]. The **NASA-TLX** questionnaire is another popular approach for measuring subjective task workload [33]. This multidimensional scale includes the rating of physical demand, which is sometimes incorrectly used to measure of fatigue. However, this dimension reflects the demands of the task itself, rather than the individual's physical response, and hence does not directly measure fatigue. In addition, many authors have raised concerns about its psychometric properties and improper use in HCI [7, 40]. Notably, even the authors of the scale themselves acknowledge that it was not designed to measure physical fatigue [33].

The **Chalder Fatigue Scale** is another tool designed to capture both physical and mental fatigue [35]. It is widely used in occupational health to quantify fatigue severity and its impact on daily functioning. Similarly, the **Multidimensional Fatigue Inventory** offers a more comprehensive assessment by evaluating multiple aspects of fatigue, including general fatigue, physical fatigue, mental fatigue, reduced motivation, and reduced activity [62]. However, one of the subscale dimensions of the Multidimensional Fatigue Inventory appears to capture anhedonia (i.e., the ability to experience pleasure or interest in activities that would typically be enjoyable [67]) instead of fatigue, raising concerns about the factor structure's validity [37]. In contrast, the **Samn-Perelli Scale** is specifically designed for use in operational settings, such as aviation, where rapid and straightforward assessments are crucial [56]. It employs a simple self-report measure that captures fatigue levels in a format that is both easy to administer and interpret. This makes it particularly useful in environments where fatigue poses immediate safety risks [53, 56, 75].

While subjective measures can provide a general estimate of fatigue, they have certain limitations. First, these measurements may struggle to detect subtle but meaningful gradations in a person's fatigue level or related constructs, as the limited number of response options restricts the granularity of the data [34, 42, 43]. Second, an individual's mental model can influence how they interpret these scales, which could potentially introduce bias [12, 40, 43]. For example, an individual's expectations about health or performance can shape how one interprets phrases like "moderate fatigue" or "severe fatigue". Third, subjective measures are typically collected retrospectively or at discrete time points, making them unsuitable for building systems that require continuous, real-time adaptation to a user's fatigue state.

Nevertheless, due to the inherently subjective nature of fatigue, self-reported ratings are commonly used to validate more objective measures. The assumption is that if both objective and subjective measures capture the same latent construct, they should be correlated. Therefore, in this study, we follow approach from prior

Table 1: The comparison of the latest shoulder fatigue models.

Models	Estimates endurance time	Does not rely on physiological sensors	Calculates recovery period	Models carried objects	Personalised arm mass calculation
CE [34]	✓	✓	✗	✗	✗
CF [36, 71]	✗	✗	✓	✗	✗
NICE [42]	✓	✓	✗	✗	✗
NICER [43]	✓	✓	✓	✗	✗
WICE (Ours)	✓	✓	✓	✓	✓

research [34, 42, 43] and employ the Borg CR10—which is a straightforward approach for categorising physical exertion—as a subjective measure of fatigue to cross-reference participants' perceptions of fatigue with the model's estimates.

2.2 Objective Fatigue Measures

The relationship between muscular exertion and fatigue has been extensively studied in different fields (e.g., sports science, ergonomics, and physiology). Methods range from external assessments, such as monitoring muscle swelling [9], muscle oxygenation [26], heart rate [10], and blood flow and pressure [61], to invasive techniques like measuring intra-arterial levels of lactate and potassium [61].

However, these approaches require specialised equipment, such as galvanic skin response (GSR) [60, 72], electrocardiogram (ECG) [45, 47, 54], electromyography (EMG) devices, and dynamometers, which limit user interaction with mid-air systems and are impractical for designing interactive systems [34]. Prior research has found a strong correlation between Borg CR10 ratings and EMG-based metrics for shoulder muscles, suggesting that either method can effectively assess shoulder fatigue [51, 68]. Additionally, studies have found the Borg CR10 scale to be more reliable than EMG metrics [34]. At low exertion levels, such as those involved in optimised mid-air gestures, EMG metrics are not valid indicators of fatigue [49]. Furthermore, EMG metrics exhibit lower repeatability compared to Borg CR10 [23], and their validity is highly task-specific [25].

Unlike these methods [8, 9, 61], models based on motion tracking as enabled through camera-based systems (e.g., OptiTrack, Microsoft Kinect, Meta Quest) provide an objective, non-invasive way to estimate shoulder fatigue. These approaches can be seamlessly integrated into interactive systems [34]. They rely on biomechanical models to estimate endurance, using inputs such as user characteristics (e.g., sex, weight) and a time series of joint angles captured by the motion tracking system (i.e., the output of the motion capture system). This is the approach that our work takes, so in the next subsection, we go through these models in detail. Table 1 shows a comparison of the characteristics of these models.

2.3 Modeling Fatigue in Mid-Air Interaction

Fatigue is a state of reduced capacity and weariness resulting from decreased endurance over time [1]. In essence, fatigue arises when the body can no longer meet the demands of continued exertion, impacting performance and endurance. This form of work-related physical fatigue is distinct from chronic fatigue linked to long-term medical conditions and is generally relieved through a period of recovery [52]. In the context of quantifying shoulder fatigue, Hincapié-Ramos et al. [34] proposed the **consumed endurance** as "a metric for characterising shoulder fatigue or gorilla-arm effects resulting from mid-air interactions". This measure is conceptual in nature and reflects the proportion of an individual's endurance capacity that has been used, combining both subjective and objective assessments. In this work, we take a similar approach, operationalising shoulder fatigue as consumed endurance, and validate using the perceived exertion (e.g., Borg CR10 rating). We did not incorporate EMG data, as the relationship between fatigue and endurance time is well established in previous fatigue models.

Endurance Time (ET) refers to the duration an individual can sustain a specific level of exertion before reaching failure or needing to rest [34, 55]. Rohmert's **Endurance Time (ET)** model [55] was the first model relevant to dynamic mid-air interactions to quantify the relationship between endurance time and shoulder fatigue using a power curve, as illustrated in Equation 1. The ET model assumes comparability between the maximum exerted force that a group of muscles can produce under controlled (*Max Force*) and the current force (*Force*) when characterising shoulder fatigue.

$$ET = \frac{1236.5}{\left(\frac{Force}{Max\ Force} \times 100 - 15\right)^{0.618}} - 72.5 \tag{1}$$

The Consumed Endurance (CE) model [34] enhances the ET model by estimating shoulder fatigue using the shoulder torque, as shown in Equation 2. Torque—the product of force and the lever arm distance—captures not only how much force is being applied but also how far from the joint that force is acting. The shoulder joint is often the pivot point during mid-air interaction. By considering the lever arm, torque calculations more accurately represent the mechanical stress on the muscles involved in holding or moving the arm in space. The cumulative average exertion level is calculated by normalising the average torque experienced on the shoulder ($\tau_{shoulder}$) with the maximum shoulder torque (τ_{max}). The CE model enables an indirect quantification of fatigue by representing physical effort as the ratio of time spent to the estimated ET, as shown in Equation 3. CE serves as a guide for VR designers in selecting interaction gestures that align with the desired fatigue level [34, 43].

$$ET = \frac{1236.5}{\left(\frac{\tau_{shoulder}}{\tau_{max}} \times 100 - 15\right)^{0.618}} - 72.5 \tag{2}$$

$$CE(T, \Delta t) = \frac{\Delta t}{ET} \times 100 \tag{3}$$

In an initial evaluation, the CE model showed a strong correlation with the Borg CR10 scale and demonstrated its potential to inform

interaction design [34]. However, the model is limited in its ability to predict the effort of low-intensity activity, to differentiate between low-intensity activity and high-intensity activity, and to predict the consumed endurance with arm positioned at angles greater than 90° [42, 43].

The original **Cumulative Fatigue (CF)** [36] and **advanced CF** [71] models improved the maximum shoulder torque estimation by using a gesture-based maximum strength model [17] based on the elbow extension angle and shoulder abduction angle. However, the CF model requires force information from the physiological sensors. Besides, CF relies on a supervised learning approach to estimate subjective fatigue ratings, which limits its ability to estimate the endurance time. Therefore, we do not consider the CF models [36, 71] in this work.

To address the limitation of CE not being able to accurately estimate the recovery period and distinguish between low and high intensity activities, the **New and Improved Consumed Endurance (NICE)** [42] and **NICE with Recovery Factor (NICER)** [43] models were developed as advancements over Rohmert's ET model by implementing empirically-derived, revised ET models. The NICE model builds upon the original CE model by incorporating static shoulder positions; however, it does not account for dynamic arm movements. The NICER model further advances this approach by introducing a shoulder-specific Endurance Time (ET) model, as shown in Equations 4 and 5. This model is derived from a meta-analysis of ET data, incorporating both static and dynamic arm gestures based on findings from prior studies [28].

$$ET = \frac{14.86}{\left(\frac{\tau_{shoulder}+C(\alpha_{shoulder})}{\tau_{max}} \cdot 100\right)^{1.83} \cdot 0.000218} \tag{4}$$

$$NICER = \frac{\Delta t}{ET} \cdot 100 \tag{5}$$

The NICER added a correction term for the shoulder torque (Equations 6 and 7) when the shoulder angle is above 90°. Besides, NICER integrates a recovery factor to account for the effect of rest periods, calculated in Equation 8. For more detailed calculations, please refer to the original papers [34, 43] and Appendix A.

$$C_{female}(\alpha_{shoulder}) = \frac{0.0095 \cdot 1005}{1+\exp\left(\frac{66.40-\alpha_{shoulder}}{7.83}\right)} - \frac{\sin\left(\alpha_{shoulder} \cdot \frac{2\pi}{360}\right)}{0.11} \quad \{90^\circ < \alpha_{shoulder} < 180^\circ\} \tag{6}$$

$$C_{male}(\alpha_{shoulder}) = \frac{0.0095 \cdot 1230}{1+\exp\left(\frac{66.40-\alpha_{shoulder}}{7.83}\right)} - \frac{\sin\left(\alpha_{shoulder} \cdot \frac{2\pi}{360}\right)}{0.09} \quad \{90^\circ < \alpha_{shoulder} < 180^\circ\} \tag{7}$$

$$NICER_{rest} = NICER \cdot \exp^{-0.04 \cdot \Delta t} \tag{8}$$

The CF and NICE models were developed based on experiments where participants held weights from 0 kg to 3 kg in their hands to refine the curve of the maximum endurance time. However, these models do not explicitly incorporate the effect of added weight into their formulas for computing consumed endurance. As a result, they are unable to accurately simulate endurance time in scenarios involving weighted objects. This limitation is significant, as virtual interactions frequently involve handheld devices such as controllers or haptic tools, which contribute to fatigue and reduced performance. In addition, Li et al. [43] capped the maximum endurance time at 5 minutes, which artificially reduced its range in the model estimation. O'Sullivan et al. [50] considered the impact of carrying objects, but their approach assumed the object's mass was evenly distributed across the limbs and focused solely on handheld items. However, in HCI applications, devices are often worn or held at specific points on the arm, concentrating weight at those locations. To address this, we propose a weighted object component within the NICER model framework that accounts for weights attached at various positions along the arm. Furthermore, all previous models for characterising shoulder fatigue use a standard average arm mass and a fixed maximum shoulder torque [27, 64]. However, such an approach fails to account for individual differences in weight and height, thereby limiting its personalisation potential. Additionally, all of these models only estimate shoulder fatigue for a bare hand (i.e., without attaching extra weights); hence, if a user is holding a controller, the model will underestimate shoulder fatigue.

To address this research gap, our goal is to capture the proportion of an individual's consumed endurance capacity without limiting the user's interaction with mid-air systems. Our model also aims to account for variations in the user's weight and height, and consider the effect of carried objects on mid-air interactions.

3 WICE: Weight-Induced Consumed Endurance Model for Mid-Air Interaction

Our work aims to produce a ready-to-use comprehensive fatigue model that accounts for the weight of objects held, carried, or worn during mid-air interactions. We investigated the strengths and limitations of existing fatigue models in the literature and summarised them in Table 1. In this section, we elaborate on the calculation of CE and NICER models (Section 3.1) and introduce new extensions: (1) a personalised arm segment mass for the calculation of the arm's centre of mass (CoM) (Section 3.2); (2) a weighted object component for determining the force resulting from the attached weighted object (Section 3.3); (3) a Bayesian regression approach for estimating endurance in a range (Section 3.4). The complete calculation of WICE is detailed in Appendix A.

3.1 Revisiting CE and NICER Models

We base the WICE model on the state-of-the-art shoulder fatigue model (NICER) as shown in Equation 4, where the calculation relies on the shoulder torque ($\tau_{shoulder}$) and the maximum shoulder torque (τ_{max}) to estimate the instantaneous exertion of the currently performed arm gesture. To determine the maximum shoulder torque (τ_{max}), the CE model uses fixed sex-specific values: 22.94Nm for males and 18.57Nm for females [27, 64]. The NICER model improves this estimate by considering the elbow extension angle and shoulder abduction angle as shown in Equation 9. This follows calculations proposed by Schanne [59] and validated by Chaffin [17]. In this equation, α_{elbow} is the elbow extension angle, and $\alpha_{shoulder}$ is the shoulder abduction angle. The τ_{max} estimate then considers sex differences through a multiplicative factor G ($G_{female} = 0.1495$, $G_{male} = 0.2845$). We adopted this value for calculation of τ_{max} in our model.

$$\tau_{max} = (227.338 + 0.525 \cdot \alpha_{elbow} - 0.296 \cdot \alpha_{shoulder}) \cdot G \tag{9}$$

Equation 10 defines the total torque ($\sum \vec{\tau}$) acting on the shoulder at a specific time. The first torque, caused by gravity ($\vec{g}$) acting on the arm's mass (M_{arm}) at the centre of mass (CoM), located at a distance ($\vec{r}$) from the shoulder joint ($\vec{r} = \overrightarrow{Sh\ CoM}$), pulls the arm downward. The second torque, generated by the shoulder muscles, counteracts gravity and facilitates arm movement. The final torque arises from the arm's inertia and angular acceleration ($\vec{\alpha}$), reflecting the arm's tendency to maintain its rotational motion once initiated.

$$\sum \vec{\tau} = \vec{r} \times (M_{arm} \cdot \vec{g}) + \vec{\tau}_{shoulder} + \vec{I}_t \times \vec{\alpha} \tag{10}$$

When the arm is static, both $\sum \vec{\tau}$ and $I\vec{\alpha}$ are zero, meaning $\|\vec{\tau}_{shoulder}\| = \|\vec{r} \times (M_{arm} \cdot \vec{g})\|$. In other words, the shoulder must generate torque equal to the gravitational torque. However, when the arm is in motion, $\sum \vec{\tau}$ and $I\vec{\alpha}$ are no longer zero. Thus, the equation for calculating the instantaneous shoulder torque at time t is shown in Equation 11.

$$\tau_{shoulder,t} = \left\| \vec{r} \times \overrightarrow{force}_{motion,t} - \left(\vec{r} \times (M_{arm} \cdot \vec{g}) + \vec{I}_t \times \vec{\alpha}_t \right) \right\| \tag{11}$$

$\vec{r}$ is the vector from the shoulder joint to the centre of mass (CoM) of the arm ($\vec{r} = \overrightarrow{Sh\ CoM}$). $\vec{g}$ is the interaction of gravity and M_{arm} is the mass of the arm at the CoM. The force acting at the CoM ($\overrightarrow{force}_{motion,t}$) depends on the acceleration ($\vec{a}_t$) at time t and the arm mass (M_{arm}) as shown in Equation 12, where the $\vec{a}_t$ can be calculated using the current arm location and time.

$$\overrightarrow{force}_{motion,t} = M_{arm} \cdot \vec{a}_t \tag{12}$$

$\vec{I}_t$ represents the moment of inertia about the shoulder axis, as defined in Equation 13. For a multi-segment structure like the arm, the total inertia is expressed as a vector whose magnitude is the sum of the inertias of each individual segment ($\vec{I}_{arm}$) [34]. The direction of this vector is determined by the unit vector ($\vec{U}_t$), which is derived from the cross product related to the movement of the centre of mass. The $\vec{U}_t$ can be calculated using Equation 14, and $\left\|\vec{I}_{arm}\right\|$ can be calculated using Equation 15, where CE used 0.0201 as the average arm inertia [27, 34, 64]. The inertia for each segment can be calculated using Equation 16, where f is the period of oscillation [27]. $\vec{\alpha}_t$ is the angular acceleration at time t as shown in Equation 17.

$$\vec{I}_t = \left\|\vec{I}_{arm}\right\| \cdot \vec{U}_t \tag{13}$$

$$\vec{U}_t = \frac{\vec{r}_{t-1} \times \overrightarrow{CoM_{t-1}\ CoM_t}}{\left\|\vec{r}_{t-1} \times \overrightarrow{CoM_{t-1}\ CoM_t}\right\|} \tag{14}$$

$$\left\|\vec{I}_{arm}\right\| = \left\|\vec{I}_{UA}\right\| + \left\|\vec{I}_{FA}\right\| + \left\|\vec{I}_{H}\right\| \tag{15}$$

$$\vec{I} = \frac{\vec{r} \times (M_{segment} \cdot \vec{g})}{4\pi^2 f^2} \tag{16}$$

$$\vec{\alpha}_t = \frac{\vec{a}_t}{\|\vec{r}\|} \tag{17}$$

3.2 Personalized Arm Center of Mass Calculation (CoM)

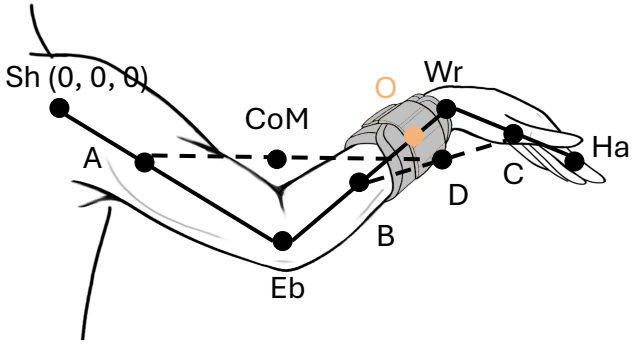


Figure 2: Arm segments that are used to calculate the centre of mass.

From Section 3.1 we can see that CoM is one of the most important part in calculating shoulder fatigue. The CoM for a two-segment structure lies along the line connecting the CoMs of each segment. Its position from the first segment's CoM is determined by the ratio of the second segment's mass to the total mass of both segments. Figure 2 illustrates the arm as a system with three segments: the upper arm (*UA* from Sh to Eb), forearm (*FA* from Eb to Wr), and hand (*H* from Wr to Ha).

Previous models use the population average as the arm mass value. However, this estimate can be improved by considering individual characteristics. Therefore, we followed Freivalds [27] to estimate the mass for each arm segment based on the individual's body mass (see Table 2 for the factors used in this calculation).

Table 2: Gross anthropometric data for the upper limbs that are used to calculate the centre of mass.

Segment	Segment mass to body mass ratio	Segment CoM to segment length ratio (from proximal end)
Upper Arm (UA)	0.029	0.452
Forearm (FA)	0.018	0.424
Hand (H)	0.008	0.397

Based on Table 2, the mass of each segment can be calculated as follows, where M_{body} is the body mass in kg:

$$M_{UA} = 0.029 \cdot M_{body} \tag{18}$$

$$M_{FA} = 0.018 \cdot M_{body} \tag{19}$$

$$M_H = 0.008 \cdot M_{body} \tag{20}$$

$$M_{arm} = M_{UA} + M_{FA} + M_H \tag{21}$$

Thus, the new CoM for each segment can be calculated as follows:

$$\vec{A} = 0.452 \cdot M_{UA} \cdot \overrightarrow{Sh\ Eb} \tag{22}$$

$$\vec{B} = 0.424 \cdot M_{FA} \cdot \overrightarrow{Eb\ Wr} \tag{23}$$

$$\vec{C} = 0.397 \cdot M_H \cdot \overrightarrow{Wr\ Ha} \tag{24}$$

$$\vec{D} = \vec{B} + \frac{M_H}{M_{FA} + M_H} \cdot \overrightarrow{BC} \tag{25}$$

$$\overrightarrow{CoM'} = \vec{A} + \frac{M_{FA} + M_H}{M_{arm}} \cdot \overrightarrow{AD} \tag{26}$$

Based on the new $\overrightarrow{CoM'}$, we can calculate a more precise $\tau_{shoulder}$ for each individual.

3.3 Weight Component Calculation (τ_{weight})

Similar to the calculation of bare-hand shoulder torque, the force exerted on the arm is equal to the gravitational force of the object, as all forces acting on the arm ultimately counteract gravity [31]. Therefore, the force can be determined using Equation 27, where $\vec{g}$ represents the gravitational acceleration at the attachment point, and M_{object} denotes the mass of the weighted object.

$$\vec{F}_{weight} = M_{object} \cdot \vec{g} \tag{27}$$

Therefore, the weighted object torque can be calculated in Equation 28, where $\vec{r}_o$ is the vector from shoulder to the weighted object's centre ($\vec{r}_o = \overrightarrow{Sh\ O}$) as shown in Figure 2. The $\vec{\tau}_{weight}$ is 0 when no weight is attached to the arm.

$$\vec{\tau}_{weight} = \vec{r}_o \times \vec{F}_{weight} \tag{28}$$

Based on the calculation of $\vec{\tau}_{weight}$, the total torque ($\sum \vec{\tau}$) acting on the shoulder with weight attached to the arm can be calculated in Equation 29. Therefore, we propose the new equation for calculating shoulder torque at time t in Equation 30.

$$\sum \vec{\tau} = \vec{\tau}_{weight} + \vec{r} \times (M_{arm} \cdot \vec{g}) + \vec{\tau}_{shoulder} + \vec{I}_t \times \vec{\alpha} \tag{29}$$

$$\tau_{shoulder,t} = \left\| \vec{r} \times \overrightarrow{force}_{motion,t} - (\vec{\tau}_{weight} + \vec{r} \times (M_{arm} \cdot \vec{g}) + \vec{I}_t \times \vec{\alpha}_t) \right\| \tag{30}$$

3.4 Bayesian Regression Model

Because we incorporate each individual's arm mass instead of using an average arm mass and revise the τ_{max} based on each individual's maximum endurance time, we obtain personalised τ_{max} and $\tau_{shoulder}$. Hence, we can model a range of the WICE values from different participants and different trials under the same condition. We employed Bayesian statistical methods in our analysis due to their enhanced flexibility, capacity to quantify uncertainty, and ability to facilitate future work to build upon it [14, 46]. This method is commonly used in HCI research [14] and studies that analysed the encumbrance effect [41]. Thus, we use a Bayesian regression model to simulate the range of WICE as shown in Equation 31, where μ is the exertion level as shown in Equation 32. Coefficients a, b, and c are provided along with their corresponding 95% compatibility intervals (see Section 5.1 for further details). The complete calculation of WICE is shown in Appendix A.

$$WICE \sim \mathcal{N}(\mu, \sigma) \tag{31}$$

$$\mu = \frac{\Delta t \cdot \left(\frac{\tau_{shoulder} + C(\alpha_{shoulder})}{\tau_{max}} \cdot 100 \right)^b \cdot c}{a} \cdot 100 \tag{32}$$

4 User Study

To validate the maximum shoulder torque (τ_{max}) for use over a wide range of motion in MR interactions, we designed a study that closely follows the design from CE [34], CF [36], and NICER [42]. In this study, we evaluated participants' endurance levels using an MR task, which was to lift a weighted object in front of the view with the required weight at the desired arm angle (Figure 3). The study was approved by our university's Human Ethics Committee.

4.1 Task and Apparatus

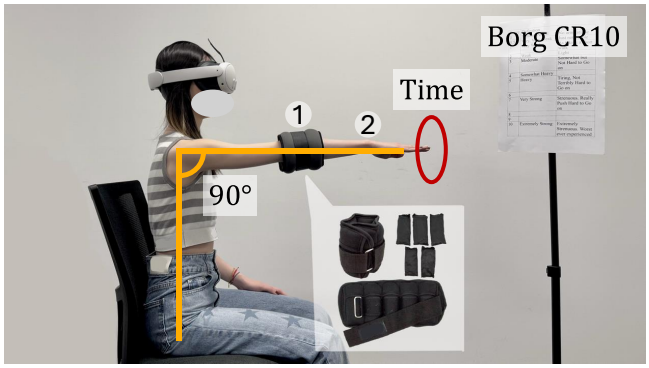


Figure 3: The experimental setup. The sandbag is attached either to location 1 (elbow) or location 2 (wrist) at a time. The red circle indicates the desired arm angle at 90° ± 5°.

The task involves participants sitting down and alternately lifting their left and right arms to a 90° position with the arm fully extended in front of the body. A sandbag is attached either to location 1 (elbow) or location 2 (wrist) during the task (Figure 3). We instructed our participants to maintain the posture until they could no longer continue. We placed the Borg CR10 scale (Appendix B) in front of the participants for their reference. During each trial, participants verbally reported their perceived fatigue every 20 seconds [36]. We used adjustable sandbags weighing 0.5 kg, 1.0 kg, 1.5 kg, and 2.0 kg, with no-weight (0 kg) condition as the baseline. These additional weight increments provided a greater number of data points, enhancing the precision of fatigue estimation compared to previous models and evaluations of encumbrance [41]. We attached the sandbag on two locations—the elbow and the wrist—as these are commonly engaged in daily activities and have been explored in previous studies [41].

We instructed participants to alternate lifting their left and right arms to allow them to fully recover between trials. We considered the engagement of both hands during mid-air interactions, rather than just the dominant hand, to replicate a realistic MR interaction scenario. Furthermore, we employ a Bayesian model to simulate the distribution of maximum endurance times across all participants. By randomly sampling measurements from both left and right hands, we capture a more comprehensive range of potential endurance limits. Consequently, our estimates are marginalised over both hands.

The arm angle between the participant's arm and their torso was set at 90°, a standard used in previous research to improve

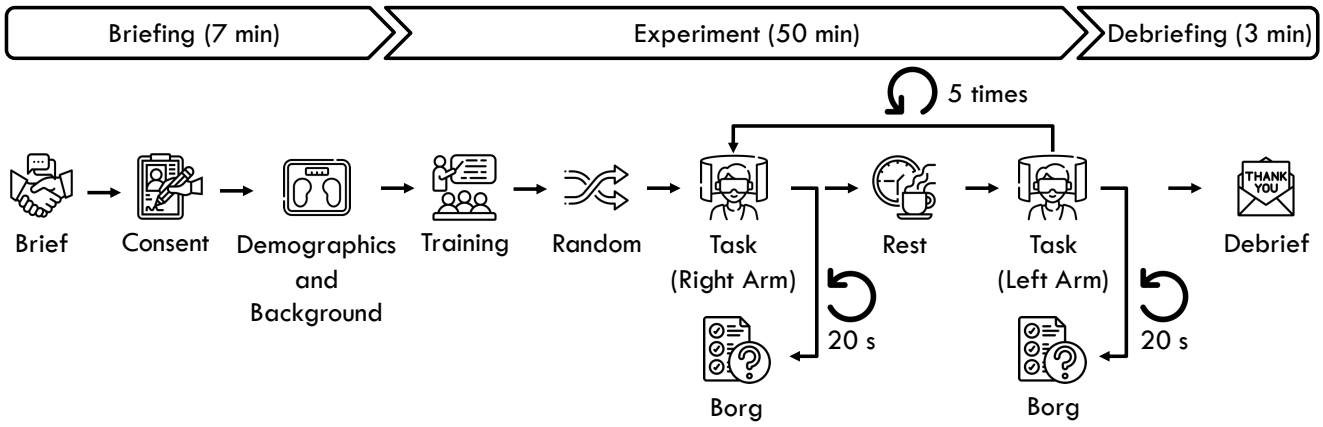


Figure 4: The experimental procedure.

the calculation of maximum shoulder torque [36]. This angle was chosen because endurance levels tend to increase gradually with higher arm angles, stabilising when the angle exceeds 90° [42, 43]. Therefore, we adopted 90° as our standard arm angle to clearly differentiate endurance levels. Participants received visual feedback in the form of a red circle and were instructed to keep their arm position within this circle to maintain it within the target range.

To measure the maximum possible duration, we set the trial duration to infinite, following previous methodologies aimed at validating maximum shoulder torque [36]. Trials concluded when participants were unable to maintain their arms within ±5° of the target angle for more than 2 seconds [36, 42].

We collected arm movement data using a Meta Quest 3 headset in a custom MR environment developed with the Unity 2022.3.58f1 engine. Participants wore the Meta Quest 3 headset, and similar to [43], we recorded arm movements using the Meta Movement SDK[5]. The location of the weighted object ($\vec{r}_o$) in Equation 28 was considered identical to the corresponding joint location.

4.2 Participant and Procedure

We recruited 38 participants (19 women, 19 men, including 3 left-handed individuals), aged between 19 and 55 years, with heights ranging from 158 cm to 186 cm, and weights from 45 kg to 103.5 kg. Our recruitment focused on an age range of 19–55—targeting current MR device users who may engage in mid-air interactions [4, 20]—which is consistent with, or even broader than, the ranges used in previous studies on shoulder fatigue [34, 36, 42, 43]. To mitigate fatigue effects, we employed a between-within (mixed) design across participants, examining two attachment locations (elbow and wrist) with four weight levels (0.5 kg, 1.0 kg, 1.5 kg, and 2.0 kg), alongside a baseline condition—without any additional weight attached. We employed a mixed-effects regression model to account for the hierarchical structure of our data, where modelling each participant as a random effect. Therefore, it is sufficient for each participant to conduct five trials. The order of conditions was randomly presented to the participants [41]. Therefore, any random effects would be distributed evenly across trials, minimising the impact of any confounding fatigue and/or learning effects.

When participants arrived at our lab, we first provided them with an overview of the study's purpose. After ensuring they understood the procedure and agreed to participate, we asked them to sign the consent form. Next, we measured each participant's weight using a body fat scale; this measurement allowed us to calculate the mass of individual arm segments accurately. To minimise any confounding effects, we asked participants to wear light clothing and remove any heavy clothing (e.g., winter jackets). Participants then completed a background questionnaire, which collected demographic details, including sex, age, and dominant hand. We gathered this information to confirm that our participant group represented a wide range of demographics and physical characteristics, thus reducing potential biases in our results.

Then, participants then performed the task according to the researcher's instructions. Between switching from one arm to the other, participants received approximately three minutes of rest. Before beginning each new trial, we checked with participants to confirm that they felt sufficiently rested. Additionally, while one arm performed the task, participants could keep the other arm relaxed. The entire study took approximately 60 minutes per participant. Figure 4 illustrates the complete experimental procedure in detail.

5 Results

We illustrated the calculation of WICE in Section 5.1. We then assessed the validity of WICE as a measure of fatigue by 1) comparing the model performance on the estimated endurance time with the ground truth endurance time on all conditions, including the baseline (Section 5.2), and 2) comparing the model performance on the estimated consumed endurance with the ground truth consumed endurance on all conditions including the baseline (Section 5.3).

5.1 WICE Model

We calculated the maximum endurance time (ET) of the WICE model using a Bayesian regression model with a shifted Log-normal distribution. In the WICE model, $\tau_{shoulder}$, τ_{max}, and $C(\alpha_{shoulder})$

[5]https://developers.meta.com/horizon/documentation/unity/move-unity-getting-started/ [Accessed: 2025-03-31]

were treated as independent variables. We included *Participant* as a random effect to account for the hierarchical structure of the data. We conducted a preliminary analysis to examine the effect of age on the experimental maximum ET. A leave-one-out cross-validation indicated that including age as a fixed effect did not enhance model performance compared to a model without age. This result suggests that age does not significantly influence ET once we accounted for other individual characteristics; therefore, we excluded it from the model's fixed parameters. For the Bayesian model, we adopted priors based on prior work, as in Equation 9. The variables included in the final model are listed below:

- $\tau_{shoulder}$: A numeric variable indicating the instantaneous shoulder torque under different conditions, which was calculated from Equation 30.
- τ_{max}: A numeric variable indicating the maximum shoulder torque under different conditions, which was calculated from Equation 9.
- $C(\alpha_{shoulder})$: A numeric variable indicating the correction term of the shoulder above 90°, calculated using Equations 6 and 7 as in NICER [43].
- *Participant*: A categorical variable used to model individual random effects.

We fit our models using the `brms` package [15], which implements Bayesian multilevel models in R using the Stan probabilistic programming language [16]. To ensure reliable results, we evaluated the convergence and stability of the Markov Chain Monte Carlo sampling using the R-hat statistic, confirming values below 1.01 [70], and verified that the Effective Sample Size (ESS) exceeded 1000 [15]. We report the posterior means of parameter estimates, the error of these estimates, and the upper and lower bounds of the 95% compatibility interval (i.e., credible interval, CI) [15]. These intervals represent the range within which there is a 95% probability that the true parameter values lie. We note that Bayesian statistics does not use p-values; therefore, our results should not be interpreted in terms of "statistical significance". The coefficients used to calculate the maximum endurance time for WICE are provided in Table 3. For full transparency, we upload all our analysis scripts and results in the supplementary material.

Table 3: Summary of the model in log scale: ET $\sim \frac{\mathbf{a}}{\left(\frac{\tau_{shoulder}+C(\alpha_{shoulder})}{\tau_{max}}\times\mathbf{100}\right)^{\mathbf{b}}\times\mathbf{c}} + (\mathbf{1}|Participant)$. We provide the posterior means of parameter estimates (Estimate), posterior error (Error), and the upper and lower bounds of their 95% CI. All parameter estimates converged with an ESS well above 1000 and an R-hat of 1.00.

	Log Scale		Original Scale	
Coef.	**Est. (Err.)**	**95% CI**	**Est. (Err.)**	**95% CI**
Multilevel Hyperparameters				
sd(b)	0.04 (0.01)	[0.03, 0.05]	1.04 (1.01)	[1.03, 1.05]
Regression Coefficients				
a	2.18 (1.15)	[0.38, 4.75]	8.84 (3.16)	[1.46, 115.66]
b	0.39 (0.03)	[0.33, 0.45]	1.48 (0.04)	[1.39, 1.57]
c	0.14 (0.07)	[0.03, 0.29]	1.15 (0.08)	[1.03, 1.34]
Further Distributional Parameters				
sigma	0.32 (0.02)	[0.28, 0.36]	1.38 (0.03)	[1.32, 1.43]
ndt	5.35 (3.19)	[0.34, 11.85]	211 (24.29)	[1.40, 140084]

5.2 Maximum Endurance Time Validation

Based on the ET model presented in Section 5.1, we evaluated the performance of the models using leave-one-out cross-validation, which provides a practical way to compare models on the basis of their predictive performance [69]. This method helps in selecting a model that is not only a good fit for the observed data but also has strong predictive power, especially when the dataset is small. It helps maximize the use of available data for both training and testing. We compared our model performance on the maximum endurance time with CE [34], NICE [42], and NICER [43].

Table 4 presents the estimated differences in expected log predictive density (ELPD Diff) along with their corresponding standard errors (SE Diff), comparing each model against WICE. The WICE model serves as the reference due to its superior performance (ELPD Diff = 0.0, SE Diff = 0.0). Lower or negative differences in expected log predictive density indicate poorer predictive performance. As can be seen from Table 4, WICE consistently outperforms all other models in estimating the maximum endurance time across all conditions. The NICER, CE, and NICE models show considerably larger negative differences in maximum endurance time (ELPD Diff = -366.3, -616.4, and -863.8, respectively), indicating substantially poorer predictive accuracy compared to the WICE model. The relatively small standard errors compared to these differences further emphasise that the performance gaps between the models are substantial. In addition, the results suggest that WICE maintains consistently strong performance across all examined conditions: female, male, weighted object, and baseline. Notably, under the baseline condition—where no additional torque from the weighted object was introduced—NICER demonstrated strong performance; however, WICE consistently outperformed the other evaluated methods using the leave-one-out cross-validation method.

We further assessed model performance by calculating the Pearson Correlation Coefficient (ρ), which measures the strength of the linear relationship between the models' posterior predictions and the observed experimental data. This metric has previously been employed to evaluate fatigue models [42, 43]. Specifically, we computed ρ between the maximum endurance times predicted by ET WICE, ET NICER, ET NICE, and ET CE, and the ground truth maximum endurance times obtained during our experiment. The coefficient ranges from -1 to 1, with values equal or greater than 0.5 indicating a strong correlation and those equal or above 0.3 suggesting a moderate correlation [22]. Table 5 shows that the WICE model achieves the highest correlation with the ground truth maximum endurance times across all conditions (overall Pearson correlation = 0.84, 95% CI = [0.80, 0.88]). In contrast, the CE, NICE, and NICER models show no meaningful correlation with the ground truth endurance times observed in our experiment.

Figure 5 illustrates the posterior predictions of maximum endurance time for each model, separated by sex and attachment location (elbow and wrist). Models whose predictions closely match the ground truth indicate better predictive accuracy. As illustrated

Table 4: The estimated difference in expected log predictive density (ELPD Diff) and their corresponding standard errors (SE Diff) when comparing the performance of each model on maximum endurance time.

	ELPD Diff (SD Diff)				
Models	**Overall**	**Female Dataset**	**Male Dataset**	**Baseline Dataset**	**Weighted Object Dataset**
ET WICE	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**
ET NICER	-366.3 (15.1)	-173.5 (12.9)	-192.9 (7.8)	-76.8 (5.1)	-289.5 (14.2)
ET CE	-616.4 (17.1)	-320.4 (14.5)	-296.0 (8.9)	-113.6 (6.6)	-502.8 (15.6)
ET NICE	-863.8 (17.1)	-423.2 (14.9)	-440.6 (8.2)	-171.2 (5.6)	-692.6 (16.1)

Table 5: Pearson correlation coefficients results (with 95% confidence intervals) between the ground truth maximum endurance time and the estimated maximum endurance time obtained from WICE, NICER, NICE, and CE across different scenarios.

	Correlation (95% CI) between ground truth ET and estimated ET				
Models	**Overall**	**Female Dataset**	**Male Dataset**	**Baseline Dataset**	**Weighted Object Dataset**
ET WICE	**0.84 (0.80, 0.88)**	**0.84 (0.77, 0.89)**	**0.83 (0.76, 0.89)**	**0.78 (0.62, 0.88)**	**0.85 (0.79, 0.89)**
ET NICER	0.14 (-0.00, 0.27)	-0.11 (-0.30, 0.10)	-0.08 (-0.28, 0.12)	0.09 (-0.23, 0.40)	0.23 (0.08, 0.38)
ET NICE	-0.15 (-0.29, -0.01)	-0.09 (-0.29, 0.11)	-0.03 (-0.23, 0.17)	-0.12 (-0.42, 0.20)	-0.21 (-0.35, -0.05)
ET CE	-0.15 (-0.29, -0.01)	-0.09 (-0.29, 0.11)	-0.02 (-0.22, 0.18)	-0.10 (-0.40, 0.23)	-0.21 (-0.36, -0.05)

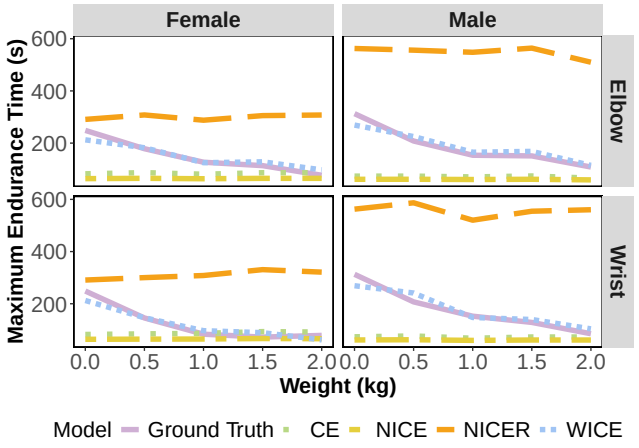


Figure 5: Comparison of endurance time (ET) predictions across different models by weight and sex, with separate panels for measurement locations (Elbow and Wrist). Models that closely align with the Ground Truth provide better estimations of maximum endurance time.

in Figure 5, the CE and NICE models underestimate the maximum ET across all experimental conditions. Moreover, these models do not adequately reflect differences between attachment locations and show minimal variation across weight levels. In particular, the NICER model exhibits substantial discrepancies at higher weights, even displaying a slight increase in predicted ET when weights are attached to female participants—a pattern inconsistent with the observed ground truth. In contrast, the WICE model provides predictions much closer to the ground truth, demonstrating superior alignment across sex, weight levels, and attachment locations. Additionally, the ground truth endurance times differ slightly depending on the attachment location, suggesting that weight impacts each location differently. The WICE model successfully captured these variations. These findings demonstrate that the WICE model provides the most reliable predictions of maximum endurance time, whereas CE, NICE, and NICER exhibit systematic biases—underestimating or overestimating endurance times.

5.3 Consumed Endurance Evaluation

Throughout this paper, we have demonstrated that WICE provides more accurate estimates of maximum endurance compared to CE, NICE, and NICER. In this section, we further evaluate each model's performance in characterising shoulder fatigue. Specifically, we assess the endurance consumed at 60 seconds—the maximum duration previously used for performance evaluation in the CE model—as this interval clearly highlights differences across various weight conditions.

We computed the Pearson Correlation Coefficient (ρ) to evaluate how closely each model's posterior predictions matched the ground truth for consumed endurance at 60 seconds (computed from the total endurance time). We compared predictions from CE, NICE, NICER, and WICE against this ground truth. The coefficient ranges from -1 to 1, with values equal to or above 0.5 indicating strong correlations and those equal to or above 0.3 suggesting moderate correlations [22]. As shown in Table 6, WICE significantly outperformed the CE, NICE, and NICER models, demonstrating a strong correlation with the ground truth consumed endurance across all conditions (overall average = 0.89, CI = [0.86, 0.92]; baseline = 0.76, CI = [0.59, 0.87]; weighted object = 0.88, CI = [0.83, 0.91]). In contrast, CE, NICE, and NICER showed no meaningful correlation between their predictions and the ground truth consumed endurance. These

Table 6: Pearson correlation coefficients (with 95% confidence intervals) between the ground truth consumed endurance at 60 seconds and the estimated consumed endurance at 60 seconds obtained from WICE, NICER, NICE, and CE across different scenarios.

	Correlation (95% CI) between ground truth CE at 60s and estimated CE at 60s				
Models	**Overall**	**Female Dataset**	**Male Dataset**	**Baseline Dataset**	**Weighted Object Dataset**
WICE	**0.89 (0.86, 0.92)**	**0.90 (0.85, 0.93)**	**0.82 (0.74, 0.88)**	**0.76 (0.59, 0.87)**	**0.88 (0.83, 0.91)**
NICER	0.18 (0.04, 0.32)	-0.16 (-0.35, 0.05)	-0.07 (-0.27, 0.13)	0.14 (-0.19, 0.43)	0.22 (0.07, 0.37)
NICE	-0.24 (-0.37, -0.10)	-0.15 (-0.34, 0.05)	-0.04 (-0.24, 0.16)	-0.07 (-0.38, 0.25)	-0.26 (-0.40, -0.10)
CE	-0.24 (-0.37, -0.10)	-0.16 (-0.35, 0.04)	-0.04 (-0.24, 0.17)	-0.06 (-0.37, 0.26)	-0.27 (-0.41, -0.11)

Table 7: Pearson correlation coefficients results (with 95% confidence intervals) between the Borg CR10 ratings at 60 seconds and the estimated consumed endurance at 60 seconds obtained from WICE, NICER, NICE, and CE across different scenarios.

	Correlation (95% CI) between Borg CR10 at 60s and estimated CE at 60s				
Models	**Overall**	**Female Dataset**	**Male Dataset**	**Baseline Dataset**	**Weighted Object Dataset**
WICE	**0.79 (0.73, 0.84)**	**0.77 (0.67, 0.84)**	**0.82 (0.74, 0.87)**	**0.65 (0.43, 0.80)**	**0.75 (0.68, 0.82)**
NICER	0.14 (-0.00, 0.28)	-0.19 (-0.37, 0.01)	-0.09 (-0.29, 0.11)	0.09 (-0.23, 0.40)	0.21 (0.05, 0.35)
NICE	-0.23 (-0.36, -0.09)	-0.18 (-0.37, 0.02)	-0.07 (-0.26, 0.14)	-0.10 (-0.40, 0.22)	-0.28 (-0.42, -0.12)
CE	-0.23 (-0.36, -0.09)	-0.18 (-0.37, 0.02)	-0.06 (-0.26, 0.14)	-0.10 (-0.40, 0.22)	-0.29 (-0.43, -0.13)

results further underline the reliability of WICE as a metric for accurately quantifying shoulder fatigue across different conditions.

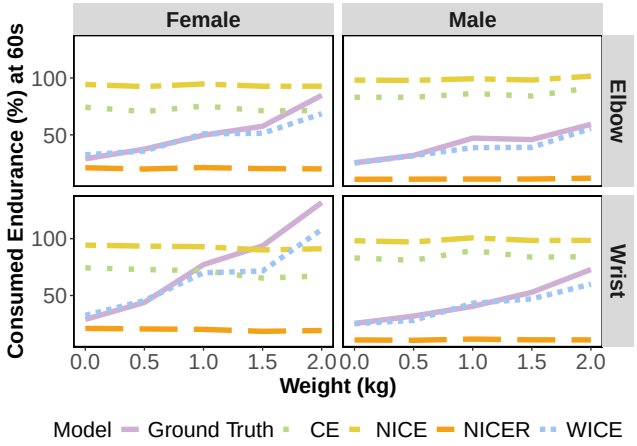


Figure 6: Comparison of consumed endurance (%) across different models by weight (kg), sex (Female and Male) and location (Elbow and Wrist). Models that closely align with the Ground Truth values provide more accurate endurance predictions.

Figure 6 shows the posterior predictions of consumed endurance (in percentage) at 60 seconds for each model, alongside the ground truth endurance observed experimentally. Models with predictions closely aligned to the ground truth demonstrate better performance. As expected, the ground truth consumed endurance increases with heavier weights, and this increase is more pronounced at the wrist than at the elbow. This suggests that attaching weight at the wrist imposes greater endurance demands and results in higher muscle fatigue compared to the elbow. The observed trends are consistent for both males and females, although males generally exhibit slightly lower consumed endurance at similar weight levels, highlighting potential sex-based differences in endurance capacity.

The CE and NICE model significantly overestimates consumed endurance, while NICER significantly underestimates consumed endurance. Additionally, as weight increases, the predicted consumed endurance at 60 seconds for CE, NICE, and NICER decreases, contradicting the observed ground truth. This inconsistency occurs because added weight causes participants' arms to lower more significantly compared to the baseline (no-weight) condition, highlighting the necessity of incorporating our proposed measure into the model. We discuss this further in Section 5.4. In contrast, both WICE models show a much closer alignment with the ground truth consumed endurance, accurately capturing the increase in endurance consumption with heavier weights and highlighting the stronger effect at the wrist compared to the elbow. Overall, WICE provides a more reliable estimate, than CE, NICE, and NICER in loaded conditions.

In addition, we also assessed the subjective feelings of fatigue at 60 seconds to validate the model's reliability. We chose Borg CR10 as the standard subjective measure of fatigue to cross-reference people's perceptions of fatigue. The Borg CR10 scale offers a ratio-scale measure of physical exertion, with its values corresponding to specific verbal descriptors [12]. The scale ranges from 0 to 10, where 0 represents "Nothing At All" and 10 signifies "Very Very Hard (Maximal)". The detailed ranges are shown in Appendix B. We chose Borg CR10 as previous research demonstrates that Borg CR10 ratings can effectively quantify shoulder fatigue [51, 68].

We calculated the Pearson Correlation coefficients for all models using the Borg CR10 ratings reported during the experiment. As can be seen in Table 7, the WICE model strongly correlated with

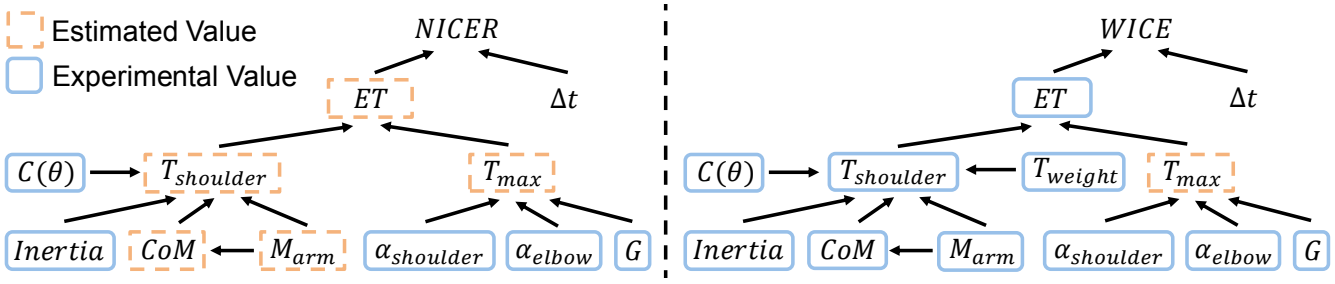


Figure 7: The comparison of the calculation of WICE and NICER.

the Borg CR10 (average = 0.79, CI = [0.73, 0.84]; baseline = 0.65, CI = [0.43, 0.80]; weighted object = 0.75, CI = [0.68, 0.82]), indicating that WICE's estimates better match users' perceptions of exertion. However, CE, NICE, and NICER showed no meaningful correlation between their estimates of consumed endurance and the ground truth.

5.4 Contrasting Model Performance

In Sections 5.2 and 5.3, we demonstrated that WICE can accurately estimate maximum endurance time and consumed endurance compared to experimental results, outperforming both CE [34], NICE [42], and NICER [43]. Figure 7 illustrates how different models estimate shoulder fatigue under various conditions. NICER serves as valuable base models, as it incorporates important factors such as joint torques (e.g., CoM and $\tau_{shoulder}$), posture angles (e.g., $\alpha_{shoulder}$ and α_{elbow}), and sex (G). This was proved through the leave-one-out cross-validation as shown in Table 4, where NICER showed a strong performance at baseline conditions. Our proposed WICE model integrates and further improves upon these factors to achieve more accurate fatigue estimation by employing a more precise measure of arm mass and incorporating a Bayesian framework.

One of the limitations of both CE and NICER is that they fail to capture the variability in arm mass across users, as they rely on an average arm mass (M_{arm}) estimated from previous research [27, 64]. Consequently, the calculation of CoM loses some individual-specific information. In contrast, our approach uses each participant's body mass to indirectly estimate individual arm mass, effectively capturing differences between users, hence accounting for inter-participant variability. Our approach estimates the mass of the arm segments based on the user's body weight, that can be easily collected from users (e.g., many exercise apps already record this information).

Furthermore, CE and NICER do not explicitly account for the effects of added weight and its position on the arm, especially when the weight is not uniformly distributed along the limb. As shown in Sections 5.2 and 5.3, instead of showing a reduction in maximum endurance time with increasing weight, CE and NICER predict either stable or even slightly increased endurance times. Similarly, for consumed endurance at 60 seconds, these models predict stable or slightly decreasing values as weight increases, directly contradicting the experimental observations.

Figure 8 explains this discrepancy: in the CE and NICER models, a lower arm position (i.e., a smaller $\alpha_{shoulder}$) is associated with a higher endurance time and, consequently, a lower consumed endurance at 60 seconds. During our user study, we observed that when the weight is attached (M_{object}) on participants' arms, participants tend to slightly lower their arms. This is because gravity drives the weighted object downward, causing the arms to fall slightly below the prescribed angle, while we allow for a margin of error of ±5° degrees. This unintended arm lowering leads both CE and NICER models to mistakenly predict an increase in maximum endurance time, contrary to actual fatigue experienced by participants. For example, Figure 9 illustrates the effect of weight placement on P33's maintained shoulder abduction angle. As the

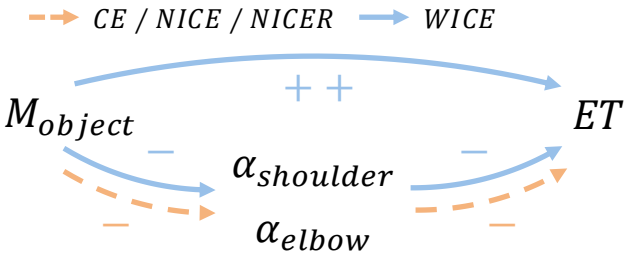


Figure 8: The effect of weight on endurance time.

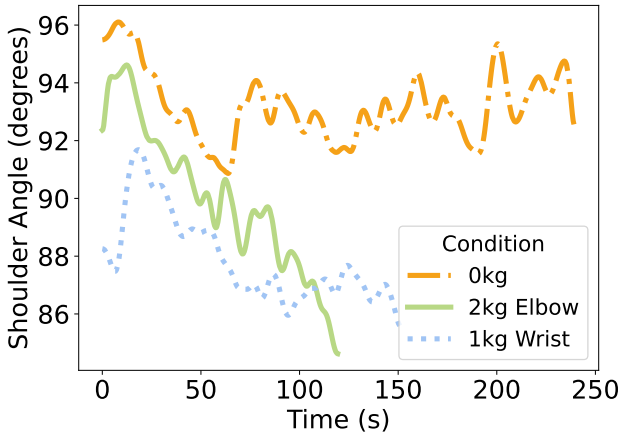


Figure 9: Shoulder abduction angle (P33) over time for three weight conditions (0 kg, 2 kg at the elbow, and 1 kg at the wrist) smoothed using a Gaussian filter.

weight increases, P33 struggles to maintain a steady arm angle, causing the arm to drop more rapidly, similar to the effect observed when the weight is positioned further from the shoulder. In practice, heavier weights consistently result in shorter maximum endurance times. Our WICE model addresses this issue by explicitly modelling the torque of the weight (τ_{weight}), thereby directly incorporating the effect of attached weight on maximum endurance time. This modification captures the impact of the attached weight on fatigue more accurately than relying solely on arm angle.

6 Discussion

Our results demonstrate that our approach effectively and accurately estimates shoulder fatigue, successfully addressing scenarios where weights are attached at various arm locations. We propose that WICE can serve as a reliable, objective measure of fatigue, offering valuable insights to guide future interaction design to design mid-air interfaces based on daily life scenarios (e.g., holding a bag while interacting in MR).

6.1 Applications of WICE

Our weight-based model offers exciting future applications in the design and evaluation of interactive systems in MR. It not only serves as an objective measure of fatigue but also offers guidance for designing more ergonomic and efficient interactions.

6.1.1 Endurance differences in using controllers vs hands-only. We employed the WICE model to estimate how much the weight of a controller can affect users' maximum endurance time. By comparing the duration a user can hold a controller in one hand across various XR headsets, our analysis quantifies the influence of controller attributes on user fatigue. We compiled a list of commonly used XR headsets with available controller weight data, gathering measurements online and verifying them with a digital scale. As shown in Table 8, the model provides posterior means, errors, and 95% credible intervals for both female and male users across different products. We used Apple Vision Pro as the baseline since it does not require controller input. The largest maximum endurance time difference observed is 17 seconds for females and 16 seconds for males, corresponding to a 7% decrease for females and a 6% decrease for males compared to no controller interaction. These metrics enable us to detect subtle performance variations resulting from differences in controller weight, offering valuable insights into how hardware modifications could enhance user comfort and operational efficiency during extended VR sessions.

Many studies are dedicated to improving VR/AR/MR experiences using haptic devices. However, an important consideration is that early prototypes for improving XR experiences tend to be clunky and significantly heavier than their eventual productions [2, 24, 38, 44, 65]. Table 9 compares several hand device prototypes and commercially available devices about the effect of weight on the maximum endurance time. For example, the commercially available device Cybergrasp[6] that can provide haptic feedback is weighted 0.450 kg. This could hinder the user's maximum endurance time by 38 seconds for females (average = 189, error = 66, 95% CI = [93, 347]) and 37 seconds for males (average = 244, error = 85, 95% CI = [119, 448]), which is around 15% decreased compared to no controller interaction. This way, researchers can use WICE to discount the effects of the additional weight inherent to the prototype and estimate endurance times based on the weight of the eventual production-grade device. This approach ensures that preliminary tests on clunky prototypes still yield insights that remain relevant once the hardware is optimised for commercial release.

[6]https://www.cyberglovesystems.com/cybergrasp [Accessed: 2025-03-31]

Table 8: The maximum duration in seconds for which a user can hold a single controller in one hand on commonly used XR headsets. We provide the posterior means of parameter estimates (Est.), posterior error (Err.), and the bounds of their 95% credible intervals (CI). W represents Weight in kg. The difference (Diff.) compared to the baseline is shown in seconds with a percentage relative to the baseline.

Product	W	Sex	Est. (Err.)	95% CI	Diff.
Apple Vision Pro	0	F	227 (80)	[111, 419]	-
		M	281 (99)	[136, 519]	-
Meta Quest 3S	0.103	F	218 (76)	[106, 400]	-9 (4%)
		M	272 (95)	[132, 501]	-9 (3%)
Meta Quest 3	0.126	F	215 (76)	[105, 397]	-12 (5%)
		M	269 (95)	[131, 497]	-11 (4%)
Meta Quest Pro	0.164	F	212 (74)	[103, 390]	-15 (7%)
		M	266 (94)	[130, 491]	-15 (5%)
HTC Vive Focus 3	0.142	F	214 (75)	[104, 394]	-13 (7%)
		M	268 (94)	[131, 494]	-13 (5%)
PlayStation VR	0.162	F	212 (74)	[103, 390]	-15 (7%)
		M	266 (94)	[130, 491]	-15 (5%)
Pico 4	0.184	F	210 (74)	[103, 387]	-17 (7%)
		M	265 (93)	[129, 487]	-16 (6%)

Table 9: The maximum duration in seconds for which a user can hold a device in one hand. We provide the posterior means of parameter estimates (Est.), posterior error (Err.), and the bounds of their 95% credible intervals (CI). W represents Weight in kg. The difference (Diff.) compared to the baseline is shown in seconds with a percentage relative to the baseline.

Product	W	Sex	Est. (Err.)	95% CI	Diff.
HapThimble [39]	0.100	F	218 (76)	[106, 401]	-9 (4%)
		M	272 (95)	[132, 501]	-9 (3%)
HandMorph [48]	0.171	F	211 (74)	[103, 389]	-16 (7%)
		M	265 (93)	[129, 489]	-16 (6%)
Dexmo [30]	0.300	F	200 (70)	[98, 368]	-27 (12%)
		M	255 (89)	[125, 470]	-26 (9%)
Cybergrasp[6]	0.450	F	189 (66)	[93, 347]	-38 (17%)
		M	244 (85)	[119, 448]	-37 (13%)
ELAXO [78]	0.540	F	183 (64)	[90, 335]	-44 (19%)
		M	237 (83)	[116, 437]	-44 (16%)
Haptic Glove [11]	0.640	F	176 (61)	[87, 323]	-51 (22%)
		M	230 (80)	[113, 423]	-51 (18%)

6.1.2 Sensing and adapting encumbrance. We argue that WICE provides a foundation for building sensing and adapting mechanisms to detect and mitigate the effects of encumbrance [41]. In practical applications, this means that systems can dynamically adjust interaction modalities (e.g., switching between gestures and speech) or re-targeting 3D objects to reduce excessive arm movement [57, 74]. Such adaptive strategies are particularly valuable in contexts like smartphones [58, 66], wearable devices [18, 76, 77], VR/AR/MR systems [41], and mid-air gaming, where maintaining user comfort is essential.

One promising direction for future development involves designing a system that leverages built-in cameras to automatically detect the object a user is holding and estimate its weight from the captured image or video [5, 63]. Such a system would continuously monitor the user's context and adapt to the corresponding solutions in real-time to maintain comfort without requiring additional effort. Future systems can deliver safer and more engaging user experiences by continuously sensing and adapting to the effects of encumbrance.

6.2 Limitations and Future Work

To manage participant fatigue and study duration, we tested only a 90° arm angle and alternated between left and right arms. By employing a between-within (mixed) design, we were able to maximise the range of conditions while ensuring a sufficient number of data points per condition. Although our study alternated tasks between the left and right arms to widen the data range and mitigate fatigue, we did not model dominance effects explicitly, where dominant hand generally displays superior motor control [6], dexterity [3], and strength [21, 73]. Future work could benefit from a focused investigation into how different arm angles, various weight locations, and handedness affect fatigue estimation.

Additionally, our model is based on the values of arm mass derived from body weight. Deriving arm mass from the participant's body mass may not accurately reflect individual variations in limb mass distribution. Nevertheless, we argue that this is currently the best method for calculating arm mass. We simplified the computation of the object's centre of mass by using the joint position, which might not capture the precise locations. This could be improved if the object is tracked individually without the need for extra tracking resources.

Finally, our model currently assumes a linear relationship between perceived exertion and maximum joint force based on previous studies [34, 36, 42, 43]. However, observations from our user study indicated that reported fatigue levels do not strictly follow a linear pattern. Indeed, some psychophysiological relationships between muscular intensity and perceived exertion are better described by power functions [13]. Moving forward, we plan to explore and identify a more accurate mapping function that captures this non-linearity to improved predictions of subjective fatigue.

7 Conclusion

This paper introduces a novel model, Weight-Induced Consumed Endurance (WICE), designed to quantify shoulder fatigue during mid-air interactions. It is the first approach to accurately characterise shoulder fatigue when additional weight is attached to different arm locations. By refining instantaneous shoulder torque calculations through the inclusion of carried object data, WICE effectively captures fatigue levels both with and without added weight. The model also incorporates individual arm mass, enabling a more personalised and accurate assessment of fatigue. Additionally, by adopting a Bayesian framework, WICE provides a probabilistic distribution of shoulder fatigue, offering richer insights compared to single-point estimates. Overall, the model demonstrated strong performance in multiple cross-validation tests, proving robust under various weight and location scenarios, and consistently outperformed existing fatigue metrics by explicitly incorporating weight-related factors. We demonstrated how WICE can be applied to examine the effects of the controller on interaction, analyse endurance differences across different levels of weight and location, provide the foundation for sensing and adapting encumbrance for more tailored MR interactions, and monitoring shoulder fatigue during interactive games.

Acknowledgments

Dr. Withana is a recipient of an Australian Research Council Discovery Early Career Award (DECRA) - DE200100479 funded by the Australian Government. We thank our participants for their valuable time. We appreciate the members of the AID-LAB for assisting us in various ways.

References

[1] Lauren S Aaronson, Cynthia S Teel, Virginia Cassmeyer, Geri B Neuberger, Leonie Pallikkathayil, Janet Pierce, Allan N Press, Phoebe D Williams, and Anita Wingate. 1999. Defining and measuring fatigue. *Image: the journal of nursing scholarship* 31, 1 (1999), 45–50.

[2] Artin Saberpour Abadian, Ata Otaran, Martin Schmitz, Marie Muehlhaus, Rishabh Dabral, Diogo Luvizon, Azumi Maekawa, Masahiko Inami, Christian Theobalt, and Jürgen Steimle. 2023. WRLKit: Computational Design of Personalized Wearable Robotic Limbs. (2023).

[3] Diane E Adamo and Anam Taufiq. 2011. Establishing hand preference: why does it matter? *Hand* 6, 3 (2011), 295–303.

[4] Robert C Allen, Michael J Singer, Daniel P McDonald, and James E Cotton. 2000. Age differences in a virtual reality entertainment environment: a field study. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 44. SAGE Publications Sage CA: Los Angeles, CA, 542–545.

[5] João Martinho Lopes Andrade and Plinio Moreno. 2023. Improving the Estimation of Object mass from images. In *2023 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, 199–206.

[6] Marian Annett. 1970. A classification of hand preference by association analysis. *British journal of psychology* 61, 3 (1970), 303–321.

[7] Ebrahim Babaei, Eduardo Velloso, Tilman Dingler, and Benjamin Tag. [n. d.]. Should We Use the Nasa-Tlx in Hci? A Review of Theoretical and Methodological Issues Around Mental Workload Measurement. *A Review of Theoretical and Methodological Issues Around Mental Workload Measurement* ([n. d.]).

[8] Myroslav Bachynskyi, Antti Oulasvirta, Gregorio Palmas, and Tino Weinkauf. 2013. Biomechanical simulation in the analysis of aimed movements. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 277–282.

[9] M Bakke, CE Thomsen, A Vilmann, K Soneda, Mauro Farella, and E Møller. 1996. Ultra sonographic assessment of the swelling of the human masseter muscle after static and dynamic activity. *Archives of oral biology* 41, 2 (1996), 133–140.

[10] Kayne Barclay, Danny Wei, Christof Lutteroth, and Robert Sheehan. 2011. A quantitative quality model for gesture based user interfaces. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference*. 31–39.

[11] Jonathan Blake and Hakan B Gurocak. 2009. Haptic glove with MR brakes for virtual reality. *IEEE/ASME Transactions On Mechatronics* 14, 5 (2009), 606–615.

[12] Gunnar Borg. 1998. *Borg's perceived exertion and pain scales*. Human kinetics.

[13] Gunnar A Borg. 1982. Psychophysical bases of perceived exertion. *Medicine and science in sports and exercise* 14, 5 (1982), 377–381.

[14] Joe Brailsford, Frank Vetere, and Eduardo Velloso. 2024. Exploring the Association between Moral Foundations and Judgements of AI Behaviour. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[15] Paul-Christian Bürkner. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80 (2017), 1–28.

[16] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of statistical software* 76 (2017).

[17] Don B Chaffin, Gunnar BJ Andersson, and Bernard J Martin. 2006. *Occupational biomechanics*. John wiley & sons.

[18] Edwin Chau, Jiakun Yu, Cagatay Goncu, and Anusha Withana. 2021. Composite line designs and accuracy measurements for tactile line tracing on touch surfaces. *Proceedings of the ACM on Human-Computer Interaction* 5, ISS (2021), 1–17.

[19] Noshaba Cheema, Laura A Frey-Law, Kourosh Naderi, Jaakko Lehtinen, Philipp Slusallek, and Perttu Hämäläinen. 2020. Predicting mid-air interaction movements and fatigue using deep reinforcement learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[20] Jiayin Chen and Calvin Or. 2017. Assessing the use of immersive virtual reality, mouse and touchscreen in pointing and dragging-and-dropping tasks among young, middle-aged and older adults. *Applied ergonomics* 65 (2017), 437–448.

[21] Anita Clerke and Jonathan Clerke. 2001. A literature review of the effect of handedness on isometric grip strength differences of the left and right hands. *The American Journal of Occupational Therapy* 55, 2 (2001), 206–211.

[22] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. routledge.

[23] Åsa Dedering, Åsa Gnospelius, and Britt Elfving. 2010. Reliability of measurements of endurance time, electromyographic fatigue and recovery, and associations to activity limitations, in patients with lumbar disc herniation. *Physiotherapy Research International* 15, 4 (2010), 189–198.

[24] Nathan Devrio and Chris Harrison. 2022. Discoband: Multiview depth-sensing smartwatch strap for hand, body and environment tracking. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

[25] Britt Elfving and Asa Dedering. 2007. Task dependency in back muscle fatigue–Correlations between two test methods. *Clinical Biomechanics* 22, 1 (2007), 28–33.

[26] Sue A Ferguson, W Gary Allread, Peter Le, Joseph D Rose, and William S Marras. 2011. Shoulder muscle oxygenation during repetitive tasks. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55. SAGE Publications Sage CA: Los Angeles, CA, 1039–1041.

[27] Andris Freivalds. 2011. *Biomechanics of the upper limbs: mechanics, modeling and musculoskeletal injuries*. CRC press.

[28] Laura A Frey Law and Keith G Avin. 2010. Endurance time is joint-specific: a modelling and meta-analysis investigation. *Ergonomics* 53, 1 (2010), 109–129.

[29] Franca Garzotto and Matteo Valoriani. 2012. " Don't touch the oven" motion-based touchless interaction with household appliances. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 721–724.

[30] Xiaochi Gu, Yifei Zhang, Weize Sun, Yuanzhe Bian, Dao Zhou, and Per Ola Kristensson. 2016. Dexmo: An inexpensive and lightweight mechanical exoskeleton for motion capture and force feedback in VR. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1991–1995.

[31] Karine Hamm. 2020. 8.6 Forces and Torques in Muscles and Joints. *Biomechanics of Human Movement* (2020).

[32] Chris Harrison, Shilpa Ramamurthy, and Scott E Hudson. 2012. On-body interaction: armed and dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. 69–76.

[33] SG Hart. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload/Elsevier* (1988).

[34] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1063–1072.

[35] Craig Jackson. 2015. The Chalder fatigue scale (CFQ 11). *Occupational medicine* 65, 1 (2015), 86–86.

[36] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 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*. 3328–3339.

[37] Leonard A Jason and Michelle Choi. 2008. Dimensions and assessment of fatigue. In *Fatigue science for human health*. Springer, 1–16.

[38] Seungwoo Je, Myung Jin Kim, Woojin Lee, Byungjoo Lee, Xing-Dong Yang, Pedro Lopes, and Andrea Bianchi. 2019. Aero-plane: A handheld force-feedback device that renders weight motion illusion on a virtual 2d plane. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 763–775.

[39] Hwan Kim, Minhwan Kim, and Woohun Lee. 2016. Hapthimble: A wearable haptic device towards usable virtual touch screen. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3694–3705.

[40] Thomas Kosch, Jakob Karolus, Johannes Zagermann, Harald Reiterer, Albrecht Schmidt, and Paweł W Woźniak. 2023. A survey on measuring cognitive workload in human-computer interaction. *Comput. Surveys* 55, 13s (2023), 1–39.

[41] Tinghui Li, Eduardo Velloso, Anusha Withana, and Zhanna Sarsenbayeva. 2025. Estimating the Effects of Encumbrance and Walking on Mixed Reality Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3706598.3713492

[42] Yi Li, Robert Crowther, Jim Smiley, Tim Dwyer, Benjamin Tag, Pourang Irani, and Barrett Ens. 2023. Revisiting Consumed Endurance: A NICE Way to Quantify Shoulder Fatigue in Virtual Reality. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*. 1–10.

[43] Yi Li, Benjamin Tag, Shaozhang Dai, Robert Crowther, Tim Dwyer, Pourang Irani, and Barrett Ens. 2024. Nicer: A new and improved consumed endurance and recovery metric to quantify muscle fatigue of mid-air interactions. *ACM Transactions on Graphics (TOG)* 43, 4 (2024), 1–14.

[44] Yu Lu, Dian Ding, Hao Pan, Yijie Li, Juntao Zhou, Yongjian Fu, Yongzhao Zhang, Yi-Chao Chen, and Guangtao Xue. 2024. Handpad: Make your hand an on-the-go writing pad via human capacitance. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

[45] M Sabarimalai Manikandan and KP Soman. 2012. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomedical signal processing and control* 7, 2 (2012), 118–128.

[46] Richard McElreath. 2018. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

[47] James P Neilson. 2015. Fetal electrocardiogram (ECG) for fetal monitoring during labour. *Cochrane database of systematic reviews* 12 (2015).

[48] Jun Nishida, Soichiro Matsuda, Hiroshi Matsui, Shan-Yuan Teng, Ziwei Liu, Kenji Suzuki, and Pedro Lopes. 2020. Handmorph: A passive exoskeleton that miniaturizes grasp. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 565–578.

[49] TOMMY Öberg, LEIF SANDSJö, and Roland Kadefors. 1994. Subjective and objective evaluation of shoulder muscle fatigue. *Ergonomics* 37, 8 (1994), 1323–1333.

[50] Patricia O'Sullivan, Matteo Menolotto, Brendan O'Flynn, and Dimitrios-Sokratis Komaris. 2025. Improving dynamic endurance time predictions for shoulder fatigue: A comparative evaluation. *Applied Ergonomics* 125 (2025), 104480.

[51] S Camille Peres, Vickie Nguyen, Philip T Kortum, Magdy Akladios, S Bart Wood, and Andrew Muddimer. 2009. Software ergonomics: relating subjective and objective measures. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. 3949–3954.

[52] Ross O Phillips. 2015. A review of definitions of fatigue–And a step towards a whole definition. *Transportation research part F: traffic psychology and behaviour* 29 (2015), 48–56.

[53] David Powell, Mick B Spencer, David Holland, Elizabeth Broadbent, and Keith J Petrie. 2007. Pilot fatigue in short-haul operations: Effects of number of sectors, duty length, and time of day. *Aviation, space, and environmental medicine* 78, 7 (2007), 698–701.

[54] Leif So Rnmo and Pablo Laguna. 2006. Electrocardiogram (ECG) signal processing. In *Wiley Encyclopedia Biomed. Eng.* 1–16.

[55] Walter Rohmert. 1960. Ermittlung von Erholungspausen für statische Arbeit des Menschen. *Internationale Zeitschrift für angewandte Physiologie einschließlich Arbeitsphysiologie* 18 (1960), 123–164.

[56] Sherwood W Samn and Layne P Perelli. 1982. *Estimating aircrew fatigue: a technique with application to airlift operations*. USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC).

[57] Zhanna Sarsenbayeva, Vassilis Kostakos, and Jorge Goncalves. 2019. Situationally-Induced Impairments and Disabilities Research. *arXiv preprint arXiv:1904.06128* (2019).

[58] Zhanna Sarsenbayeva, Niels Van Berkel, Aku Visuri, Sirkka Rissanen, Hannu Rintamaki, Vassilis Kostakos, and Jorge Goncalves. 2017. Sensing cold-induced situational impairments in mobile interaction using battery temperature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–9.

[59] Frederick John Schanne Jr. 1972. *A Three-Dimensional Hand Force Capability Model for a Seated Person.(Volumes I And II)*. University of Michigan.

[60] Yu Shi, Natalie Ruiz, Ronnie Taib, Eric Choi, and Fang Chen. 2007. Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*. 2651–2656.

[61] Gisela Sjøgaard, Gabrielle Savard, and Carsten Juel. 1988. Muscle blood flow during isometric activity and its relation to muscle fatigue. *European journal of applied physiology and occupational physiology* 57 (1988), 327–335.

[62] EMA Smets, Bert Garssen, B de Bonke, and JCJM De Haes. 1995. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of psychosomatic research* 39, 3 (1995), 315–325.

[63] Trevor Standley, Ozan Sener, Dawn Chen, and Silvio Savarese. 2017. image2mass: Estimating the mass of an object from its image. In *Conference on Robot Learning*. PMLR, 324–333.

[64] Hong Z Tan, Mandayam A Srinivasan, Brian Eberman, and Belinda Cheng. 1994. Human factors for the design of force-reflecting haptic interfaces. *Dynamic Systems and Control* 55, 1 (1994), 353–359.

[65] Shan-Yuan Teng, KD Wu, Jacqueline Chen, and Pedro Lopes. 2022. Prolonging VR haptic experiences by harvesting kinetic energy from the user. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–18.

[66] Garreth W Tigwell, Zhanna Sarsenbayeva, Benjamin M Gorman, David R Flatla, Jorge Goncalves, Yeliz Yesilada, and Jacob O Wobbrock. 2019. Addressing the challenges of situationally-induced impairments and disabilities in mobile interaction. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

[67] Michael T Treadway and David H Zald. 2011. Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neuroscience & Biobehavioral Reviews* 35, 3 (2011), 537–555.

[68] Amedeo Troiano, Francesco Naddeo, Erik Sosso, Gianfranco Camarota, Roberto Merletti, and Luca Mesin. 2008. 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.

[69] Aki Vehtari, Andrew Gelman, and Jonah Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing* 27 (2017), 1413–1432.

[70] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. 2021. Rank-normalization, folding, and localization: An improved R^ for assessing convergence of MCMC (with discussion). *Bayesian analysis* 16, 2 (2021), 667–718.

[71] Ana Villanueva, Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2023. Advanced modeling method for quantifying cumulative subjective fatigue in mid-air interaction. *International Journal of Human-Computer Studies* 169 (2023), 102931.

[72] María Viqueira Villarejo, Begoña García Zapirain, and Amaia Méndez Zorrilla. 2012. A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee. *Sensors* 12, 5 (2012), 6075–6101.

[73] Michael WB Watterworth, Fahima Wakeely, Sarah A Fitzgerald, and Nicholas J La Delfa. 2024. The effect of handedness on upper extremity isometric strength symmetry. *Applied Ergonomics* 114 (2024), 104133.

[74] Jacob O Wobbrock. 2019. Situationally-induced impairments and disabilities. *Web Accessibility: A Foundation for Research* (2019), 59–92.

[75] Alexander Wolkow, Brad Aisbett, John Reynolds, Sally A Ferguson, and Luana C Main. 2016. Acute psychophysiological relationships between mood, inflammatory and cortisol changes in response to simulated physical firefighting work and sleep restriction. *Applied psychophysiology and biofeedback* 41 (2016), 165–180.

[76] Jiakun Yu, Supun Kuruppu, Biyon Fernando, Praneeth Bimsara Perera, Yuta Sugiura, Sriram Subramanian, and Anusha Withana. 2024. IrOnTex: Using Ironable 3D Printed Objects to Fabricate and Prototype Customizable Interactive Textiles. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8, 3 (2024), 1–26.

[77] Jiakun Yu, Praneeth Bimsara Perera, Rahal Viddusha Perera, Mohammad Mirkhalaf Valashani, and Anusha Withana. 2024. Fabricating customizable 3-D printed pressure sensors by tuning infill characteristics. *IEEE Sensors Journal* 24, 6 (2024), 7604–7613.

[78] Zhong-Yi Zhang, Hong-Xian Chen, Shih-Hao Wang, and Hsin-Ruey Tsai. 2022. ELAXO: Rendering Versatile Resistive force feedback for fingers grasping and twisting. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14.