Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-Air Interactions

Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, Pourang Irani
University of Manitoba
Winnipeg, MB, Canada
{hincapjd, umguo62, umpaymah, irani}@cc.umanitoba.ca

ABSTRACT
Mid-air interactions are prone to fatigue and lead to a feeling of heaviness in the upper limbs, a condition casually termed as the gorilla-arm effect. Designers have often associated limitations of their mid-air interactions with arm fatigue, but do not possess a quantitative method to assess and therefore mitigate it. In this paper we propose a novel metric, *Consumed Endurance* (CE), derived from the biomechanical structure of the upper arm and aimed at characterizing the gorilla-arm effect. We present a method to capture CE in a non-intrusive manner using an off-the-shelf camera-based skeleton tracking system, and demonstrate that CE correlates strongly with the Borg CR10 scale of perceived exertion. We show how designers can use CE as a complementary metric for evaluating existing and designing novel mid-air interactions, including tasks with repetitive input such as mid-air text-entry. Finally, we propose a series of guidelines for the design of fatigue-efficient mid-air interfaces.

Author Keywords
Gorilla-arm, mid-air interactions, mid-air text-entry, endurance, consumed endurance, SEATO mid-air keyboard.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): Evaluation/Methodology.

INTRODUCTION
The proliferation of low-cost gestural tracking systems has warranted the investigation of mid-air interaction as a new class of natural user interface (NUI) [20, 19]. This style of interaction has shown particular value in sterile medical rooms [7, 25], in educational settings [12], and in gaming environments [22]. Nonetheless, users engaged with mid-air input often report fatigue and a feeling of heaviness in the arm [9, 20], a condition coined as the gorilla-arm effect [9]. Gorilla-arm was first reported with the introduction of touch-screens, and was one reason for the early dismissal of such systems [1, 2]. Ignoring this factor in the design of mid-air interactions can also lead to the demise of this form of NUI.

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada
Copyright 2014 ACM 978-1-4503-2473-1/14/04…$15.00.
http://dx.doi.org/10.1145/2556288.2557130

Current approaches to assess arm fatigue include obtrusive measurements of bodily variables (heart-rate [30], oxygen level [16] or EMG [28, 32]) or the collection of subjective assessments (Borg [8], NASA-TLX [21] or Likert ratings). However, these methods have limited practical value for evaluating mid-air interactions as they require specialized equipment or have high variance. We propose a method for quantitatively characterizing the gorilla-arm effect based on the concept of endurance (Figure 1 and equation 1) [29, 17]. Endurance is the amount of time a muscle can maintain a given contraction level before needing rest. Using a skeleton-tracking system we capture users' arm motions and compute endurance for the shoulder muscles. *Consumed Endurance* (CE), our novel metric, is the ratio of the interaction time and the computed endurance time.

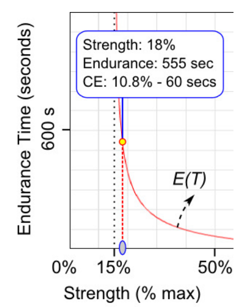

Figure 1. Endurance metrics for the mid-air interactions include Arm Strength, Endurance and Consumed Endurance (CE).

We validate CE against fatigue ratings as obtained using the Borg CR10 scale of perceived exertion. Further, we demonstrate CE's value as a complementary metric for evaluating mid-air interactions. For mid-air pointing and selection on a 2D plane, we used CE to identify the most suitable interaction parameters, such as arm extension, plane location and plane size. For example, users consumed the least amount of endurance when the arm was bent and operating on the interaction plane located midway between the shoulder and the waist. *Dwell* selections have the lowest CE for single hand interactions. We also demonstrate the value of using CE to inform the design of an endurance-efficient text-entry layout, SEATO (Figure 7-left). Users entering text with SEATO had lower CE than with QWERTY without compromising text-entry speed. Finally, we describe how our results inform the design of mid-air menus and other interactive systems. CE and other fatigue-related metrics are publicly available in a software toolkit (http://hci.cs.umanitoba.ca/projects-and-research/details/ce).

Our contributions include: 1) CE, a metric for characterizing shoulder fatigue or gorilla-arm effects resulting from mid-air interactions; 2) the use of CE to inform the choice of various mid-air interaction parameters; 3) an endurance-efficient mid-air text-entry layout, SEATO; and 4) guidelines for designing endurance-efficient mid-air interactions.

RELATED WORK

This section gives a summary of the existing qualitative and quantitative tools for assessing muscular fatigue.

Qualitative Assessment of Fatigue

Qualitative methods for assessing arm fatigue include Likert-scale questions [9], the NASA Task Load Index (NASA-TLX) [21] and the Borg RPE and CR10 scales [8]. Likert-scales reduce the users' subjective ratings to whether or not they experienced fatigue in an interaction. The NASA-TLX questionnaire captures workload along categories such as Physical Demand and Effort [21], rated on a 20-point scale. However, as pointed by Bustamante and Spain, the TLX lacks "scalar invariance, thereby biasing the estimation of mean scores and making the examination of mean differences misleading" [10]. Subjectivity is further reinforced as each participant can weigh the various TLX categories differently. The Borg CR10 scale [8] is tailored to physical exertion. It maps numeric ratings to carefully chosen verbal cues and provides scalar invariance.

While qualitative assessments provide a coarse estimation of fatigue, a finer characterization is required, particularly for repetitive tasks. Subjective assessments cannot give an account of the small yet significant differences, and are prone to confounding variables such as the participant's fitness, comfort level or general state of mind. Complementing such methods with objective metrics of fatigue can provide a more holistic handle over gorilla-arm effects.

Objective Assessment of Fatigue

Fields such as the sports sciences, ergonomics and physiology have long studied the relationship between muscular exertion and fatigue. Their methods range from external measurements such as monitoring muscle swelling [5], muscle oxygenation [16], heart rate (mentioned in [6]), and blood flow and pressure [30]; to invasive techniques such as measuring the intra-arterial levels of lactate and potassium [30]. Morris et al. [26] show a strong relationship between different fatigue-related factors obtained externally (rate of force exertion and relaxation) and those captured invasively (lactate, oxygen level), indicating that the former are highly reliable measures of fatigue. However, these approaches require specialized equipment (such as EMG and NIRS devices, dynamometers or exoskeletons) or are invasive, limiting how users engage with mid-air systems, and thus impractical for the design of interactive systems. Unlike previous approaches [5, 16, 30, 4], our method provides an objective and non-invasive metric of shoulder fatigue, calculated using a low-cost gestural tracking system.

FATIGUE IN MID-AIR INTERACTIONS

Gorilla-arm is a manifestation of fatigue in the arm muscles. Fatigue is defined as the ability to maintain a given muscular contraction level [17] and depends on the amount of blood flow, and thus oxygen, that reaches the muscle cells. A contracted muscle hardens arteries and restricts blood flow. With low levels of oxygen muscle cells switch their energy source from aerobic to glycolytic metabolism. Given the limited amounts of stored glucose, the muscle cells can produce energy and maintain the contraction only for a short period of time. Fatigue occurs when this energy is used up.

A well accepted result in human physiology is Rohmert's study of the impact of fatigue on endurance: the maximum amount of time that a muscle can maintain a contraction level before needing rest [29]. Figure 1 illustrates Rohmert's formulation of endurance *E(F)* as a function of the value of force applied (F) in relation to the maximum force (F_{max}) of the muscle (see equation 1). An important observation is that equation 1 is asymptotic at 15% of the maximum force, meaning that forces exerted below that level could be sustained for long time periods. The presence of fatigue in mid-air interactions [9, 20] suggests that current interaction techniques require arm forces above the 15% mark.

$$E(F) = \frac{1236.5}{\left(\frac{F}{F_{max}} * 100 - 15\right)^{0.618}} - 72.5 \qquad (1)$$

We derive metrics based on this mathematical formulation of endurance to study and guide the design of mid-air interact-tions. A higher endurance time for an interaction implies that it triggers lower amounts of fatigue in users and thus allows for longer engagements with a system. Ultimately, this should foster a broader adoption of such technologies.

CHARACTERIZING SHOULDER ENDURANCE

This section details a non-intrusive method for determining endurance for mid-air interactions. Although multiple body parts are involved in mid-air arm interactions we focus on the shoulder joint as it largely dominates the forces required for moving the arm. Measuring endurance using Rohmert's formulation requires capturing the two variables in equation 1: the maximum force of the shoulder (F_{max}), and the force acting on the shoulder at a given time (F). For the first variable, we rely on values of maximum force as determined previously by others [31, 14]. For the second variable, we use a biomechanical model of the arm where it is represented as a compound rotational system encompassing the upper arm, the forearm and the hand, and with its pivot on the shoulder joint (see Figure 2-left). This compound system can be simplified as a single-part system where all forces are applied at the arm's center of mass (CoM) (see Figure 2-right).

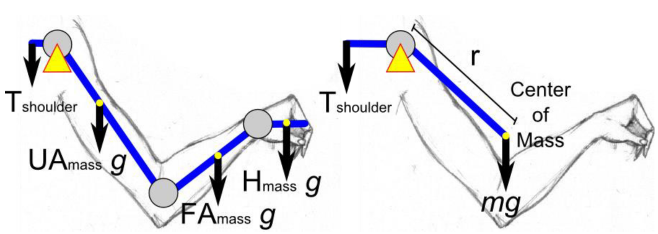

Figure 2. Left: The primary forces acting on the arm. Right: The forces aggregated at the CoM.

Rohmert's formulation assumes that F and F_{max} are com-parable, that is, they are applied at the same distance from the shoulder joint. However, given that F_{max} is determined at the elbow [31] and F at the CoM, these two forces are not comparable. A solution is to express all forces in terms of torque. Appendix A shows the relationship between $\frac{\|\vec{F}\|}{F_{max}}$ and $\frac{\|\vec{T}\|}{T_{max}}$; and therefore endurance in terms of torque is:

$$E(T_{shoulder}) = \frac{1236.5}{\left(\frac{T_{shoulder}}{T_{max}} * 100 - 15\right)^{0.618}} - 72.5 \qquad (2)$$

Equation 3 formalizes the sum of torques ($\sum \vec{T}$) acting on the system at a given time. The first torque pulls the arm downwards and is due to the interaction of gravity (g) and the mass of the arm (m) at the CoM (distance r from the shoulder joint). The second torque is provided by the shoulder muscles and it compensates for the effects of gravity and moves the arm. The final torque is due to the arm's inertia and its angular acceleration ($I\alpha$), and represents the tendency of the arm to maintain its rotational movement once in motion.

$$\sum \vec{T} = \vec{r} \times m\vec{g} + \vec{T}_{shoulder} + \overrightarrow{I\alpha} \qquad (3)$$

When the arm is static the $\sum \vec{T}$ and $\vec{I}\alpha$ components are equal to zero, and therefore $\|\vec{T}_{shoulder}\| = \|\vec{r} \times m\vec{g}\|$. That is, the shoulder has to match the gravity torque. Conversely, when the arm is in rotation the resulting $\sum \vec{T}$ and $\overrightarrow{I\alpha}$ are not equal to zero. The next section shows how to calculate $\vec{T}_{shoulder}$.

Measuring Torque at the Shoulder Joint

By following the process described in Appendix B, a skeleton tracking system can compute the CoM location r at time t. Tracking the CoM allows us to determine its velocity and acceleration. Knowing the arm mass it is possible to determine the force acting at the CoM:

$$\overrightarrow{force}_{motion,t} = \overrightarrow{acceleration}_t * m \qquad (4)$$

Thus, the total torque of the system can be expressed as[1]:

$$\sum \vec{T}_t = \vec{r} \times \overrightarrow{force}_{motion,t} \qquad (5)$$

Equaling equations 3 and 5, we can derive the actual torque exerted by the shoulder muscles at time t:

$$\|\vec{T}_{shoulder,t}\| = \|\vec{r} \times \overrightarrow{force}_{motion,t} - (\vec{r} \times m\vec{g} + I_t \vec{\propto}_t)\| \qquad (6)$$

The final term in equation 6, $I_t \vec{\propto}_t$, represents the tendency of a rotating body to continue its rotational movement. Angular acceleration ($\vec{\propto}$) at time t can be calculated as $\vec{\propto}_t = \frac{acceleration_t}{\|\vec{r}\|}$. See Appendix C for a detailed account on how to calculate the moment of inertia (I) of a moving arm.

Maximum Torque of the Shoulder

Tan et al. studied the force output range at different joints in the body [31], and established the maximum controllable force of the shoulder applied at the elbow to be 87.2 *N* for females and 101.6 *N* for males. In a similar study, Edmunds et al. found that such maximum force varies slightly according to the movement direction (x,y,z), with an approximate 100 *N* as the most common value [14]. We use Tan et al.'s maximum force to get the maximum torque T_{max} of the shoulder muscles; based on their experimental set-up, we remove the effect of the arm weight from the max torque.

Endurance Metrics

Based on the above, we can calculate the following metrics:

- Strength (S): The ratio between the average torque applied in the interaction and the maximum torque. This metric corresponds to the $\frac{\|\vec{T}_{shoulder}\|}{T_{max}} * 100$ term of equation 5.
- Endurance (E): The time, in seconds, the participant could sustain such interaction before needing to rest the arm.
- Consumed Endurance (CE): The ratio of the interaction time and the computed endurance time (see equation 7). We interpret CE as the percentage of the energy used or as the amount of fatigue.

$$CE(T, TotalTime) = \frac{TotalTime}{E(T)} * 100 \qquad (7)$$

ENDURANCE METRICS IMPLEMENTATION

We used Microsoft's .NET to implement the CE equations. We used the Microsoft Kinect, an off-the-shelf human skeleton tracking system, to capture the arm joints needed. We applied a noise reduction filter averaging the last 10 skeleton frames. The system operates at approximately 32 frames/sec, which is sufficient to support the small delta (∂t) assumption made for equation 5. We used the 50th percentile male or female values[2] for weight, length, center of mass, and inertia of the upper arm, lower arm, and hand (a standard approach [18]) as compiled by Freivalds [17]. Using these values and the captured skeleton, our system determines the arm's CoM and normalizes it to the 50th percentile (we use the skeleton's upper arm's length as a frame of reference). The system calculates all metrics from the normalized CoM.

VALIDATING THE CONSUMED ENDURANCE METRIC

In this section, we assess the validity of CE as a measure of fatigue by 1) comparing CE measurements to Borg CR10 ratings [8] and 2) analyzing the effects of gender-specific constants. The Borg CR10 scale provides a ratio-scale measure of physical exertion which values are matched to verbal anchors. Borg CR10 values range from 0 to 10, where 0 corresponds to "Nothing At All" and 10 to "Very Very Hard (Maximal)". Electromyograms (EMG) measure muscle cell activations and is the bases for several objective metrics (see [32]). Studies have shown that Borg CR10 ratings and EMG-based metrics for shoulder muscles strongly correlate and therefore either method can be used to assess shoulder fatigue [28, 32]. More importantly, they showed that Borg CR10 can be more reliable than EMG metrics. At low levels of physical exertion (such as the ones in optimized mid-air gestures) EMG metrics are not valid fatigue indicators [27]. Also, EMG metrics have lower repeatability than Borg CR10 [13] and their validity is task-dependent [15].

We asked 16 participants (8 female) to hold their dominant arm at different angles from the vertical axis (90°, 60°, 30°,

[1] The cross product in equation 5 indicates that only the forces tangential to the rotation axis are relevant for the torque. When ($\partial t \to 0$), then $\overrightarrow{force}_{motion,t} \to \overrightarrow{force}_{tangential,t}$.

[2] An alternative method measures the length of a volunteer's upper limbs and uses this data to retrieve the average corresponding weights based on data provided by the Visible Human Project [3].

and at rest, controlled within ±2 degrees) and for different periods of time (15, 30, 45 and 60 seconds). We used a Latin-square design on *angle* with *time* and *gender* as random factors. We captured 3 trials per condition per participant for a total of 16×16×3 = 768 CE measurements. Participants rated each condition (3 trials) for a total of 16×16 = 256 Borg CR10 ratings. Figure 3 shows the overall results.

We used linear regression analysis and the Mixed Factors ANOVA test with *angle* and *time* as within-subject factors and *gender* as a between-subjects factor. We used Bonferroni correction for post-hoc tests. Linear regression analysis between CE and Borg CR10 ratings (see Figure 3-bottom) revealed a significant correlation ($F_1 = 1902.722$, $p < 0.001$) with $R^2 = 0.716$. For the ANOVA, normality tests showed a normal distribution for all conditions ($p < 0.001$). Results showed a main effect in *angle* for CE ($F_{3,42} = 8543.719$, $p < 0.001$) and Borg CR10 ($F_{3,42} = 89.806$, $p < 0.001$); a main effect in *time* for CE ($F_{3,42} = 23323.431$, $p < 0.001$) and Borg CR10 ($F_{3,42} = 68.460$, $p < 0.001$); and no main effect in *gender* for CE ($F_{1,14} = 0.951$, $p = 0.346$) and Borg CR10 ($F_{1,14} = 2.379$, $p = 0.145$). Results showed interaction effects for *angle* × *time* in CE ($F_{9,126} = 4874.636$, $p < 0.001$) and Borg CR10 ($F_{9,126} = 10.301$, $p < 0.05$). Post-hoc analysis of *angle* and *time* showed significant differences between all conditions for both CE and Borg CR10.

Our results show that CE and Borg CR10 ratings present a "very strong to perfect association" ($R = 0.846$) where the value of CE is used to predict 72% of the variability of Borg CR10 ratings ($R^2 = 0.716$). The remaining 28% can be explained by differences in fitness level of the participants and the subjective nature of the Borg CR10 scale (affected by factors such as tiredness, comfort, and general state of mind). Moreover, CE and Borg CR10 ratings are equally capable of yielding significant differences for changes in *angle* and *time* (main effects on both factors). Furthermore, results show that CE is gender neutral, suggesting that the different sets of constants for arm metrics (weights, lengths, and max force) do not affect CE. In other words, given Borg CR10's correlation to objective measurements of fatigue (such as EMG), these results show that CE is a valid objective fatigue metric for the shoulder muscles.

CONSUMED ENDURANCE AS AN ANALYTICAL TOOL

We first demonstrate the use of our model to evaluate different mid-air interaction factors. In the first experiment we investigate the effects of different plane locations and arm extensions on CE. In the second experiment we study the effects of plane size and selection method on CE.

Experiment 1: Plane Location and Arm Extension

Methods

Apparatus – The system ran on a Windows 7 PC connected to a 4×2.3 meters projector screen with a resolution of 1366× 768 pixels (1 pixel = 3 mm) and a Microsoft Kinect. The Kinect was in front of the screen and 1 meter above the floor; participants stood 3.3 meters from the screen. We used the same set-up for all experiments.

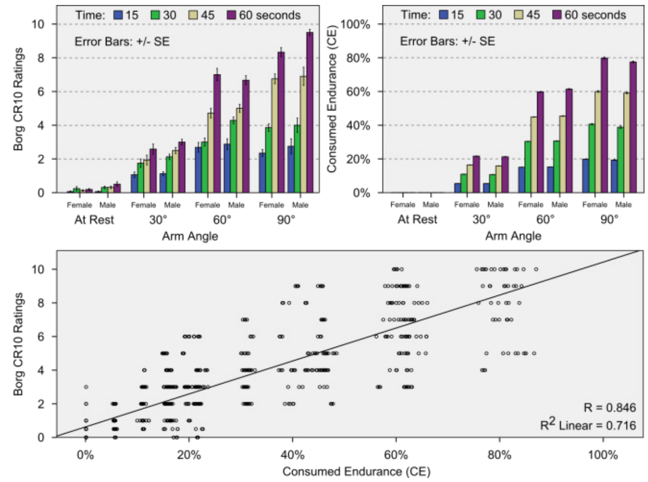

Figure 3. Top – Both CE and Borg CR10 present a main effect for *angle* and *time* but not for *gender*. Bottom – Linear correlation between Borg CR10 and CE show a strong correlation (R = 0.846) where CE predicts 72% of the variability in Borg CR10 ratings (R^2 = 0.716).

Subjects – 12 participants (3 female) volunteered, ages 18-40 (mean 26), right handed. All participants had previous experience (mean: 0.6 years) with mid-air interaction systems (Wii, Kinect, etc.) and were familiar with mid-air selection.

Task – Participants had to select 20 fixed targets (one after the other) in a square 2D plane (35 cm sides) by moving the cursor (small red circle) with their right arm from the current position to the target. Participants were asked to select using a mouse button held in the left hand. We relegated selection to a mouse to avoid any overhead. All targets were solid squares organized in a 6×6 matrix (black border, white background). Upon selection, the target was highlighted in red, and the next target turned blue. The task finished when the participant selected the 20 targets in the order presented. The 20 targets were randomly distributed across all positions and no position was repeated. A landing error was marked when the user left the target before selecting it. This measure describes the level of control a user has over the cursor, i.e., how precise the movements are.

Design – Independent variables were *plane location* and *arm extension* (see Figure 4). We used a 2×2 within-subject design to compare CE in each condition. We considered two 2D plane locations relative to the body (all planes aligned to the right side of the shoulder):

- *Shoulder*: is a vertical plane with the vertical center at the shoulder joint.
- *Center*: is a vertical plane with the vertical center located halfway between the shoulder and the waist.

We considered two arm extensions: Extended and Bent. The system detects the arm as extended when the hand is at least

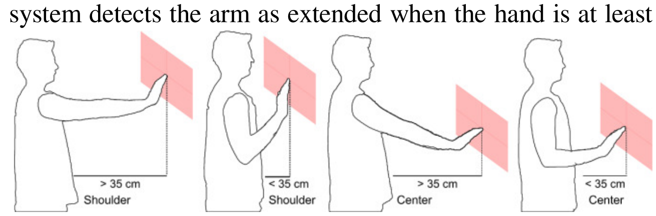

Figure 4. Two location planes (shoulder, center) with two arm configurations (extended, bent)

35 cm away from the body, and as bent when the hand is 35 cm or closer to the body plane. The system ignores the arm (removes the cursor from the screen) when it is under or beyond these limits, forcing the participant to stretch out or bend it as necessary. We settled on these measures after iterative pilot testing.

Participants were trained with each condition after the experimenter demonstrated the task. With a total of 2×2 = 4 conditions and 4 trials per condition, we registered 2×2×4 = 16 trials per participant, or 192 trials in total (each trial consisted of 20 selections). Participants had a mandatory 3 minute break between conditions. All participants completed the experiment in one session lasting approximately 30 minutes. The trials were counter-balanced using a Latin-square on plane location and arm extension.

Measures –We collected values for CE, completion time, and landing error rate. Participants filled a Borg CR10 rate scale questionnaire after each condition.

Results

None of the dependent variables comply with the ANOVA assumptions (normality and equal variances) and therefore we applied the Aligned Rank Transform for nonparametric factorial analysis [33] with a Bonferroni correction for pairwise comparisons. Figure 5 presents the results.

Consumed Endurance (CE) – Results showed a main effect of *plane location* ($F_{1,11}$ = 102.249, p < 0.001) and *arm extension* ($F_{1,11}$ = 86.959, p < 0.001). There were not significant interaction effects for *plane location* × *arm extension* (p = 0.637). CE was lowest in the center plane location with an average of 27.23% (standard deviation or std = 15.33) and the bent arm extension at 23.44% (std = 13.45).

Borg CR10 – Results showed main effects of *plane location* ($F_{1,11}$ = 7.111, p < 0.05) and *arm extension* ($F_{1,11}$ = 21.082, p < .001). There were no significant interaction effects for *plane location* × *arm extension* (p = 0.134). Borg CR10 was lowest on the center plane location at 3.00 (std = 1.87) and the bent arm extension at 2.75 (std = 1.51).

Completion Time – Results did not show a main effect of *plane location* (p = .092) or *arm extension* (p = .223). There were no significant interaction effects for *plane location* × *arm extension* (p = .893). Average completion time was 45 seconds (std = 13.81).

Landing Error Rate – Results did not show a main effect of *arm extension* (p = .619) or *plane location* (p = .357). There were no significant interaction effects for *plane location* × *arm extension* (p = .220). Average landing error was 0.93 (std = 2.22).

Discussion

We first observe that both CE and Borg CR10 yield similar main and interaction effects, highlighting CE's capacity to reveal the same fatigue effects as Borg CR10. On the other hand, differences in completion times and error rates are not significant. This is an important observation because it suggests that differences in fatigue emerge even when other measurements are flat. Therefore, completion time and

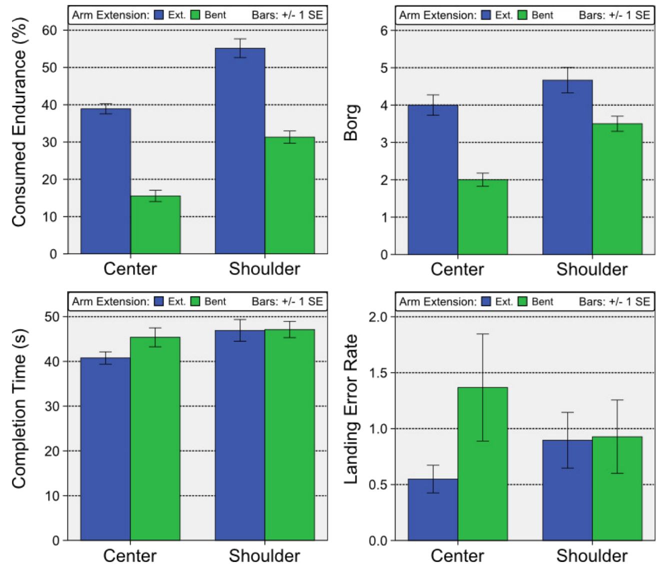

Figure 5. Consumed Endurance, Borg CR10 ratings, completion time and landing error rate for experiment one.

landing error rate were limited in determining the optimal combination of *plane location* and *arm extension* for mid-air input, given our conditions. With equivalent accuracy to Borg CR10, a system can calculate CE unobtrusively and in real-time by simply tracking arm movements.

Interactions in the shoulder plane consume more endurance as the arm is higher up from its resting position. Similarly, interactions with arm extended also consumed more endurance as the center-of-mass is further extended from the body, thus requiring a higher torque. Interactions with the bent arm consumed the least endurance, the lowest being in the center plane at 15.55%. We select the center + bent condition as the optimal area for interaction and use it in the next experiments which evaluate CE for other factors.

Experiment 2 - Plane Size and Selection Method

The goal of this experiment is to examine the effect of larger arm movements and different selection methods on CE.

Methods

Subjects – 12 participants (4 female), ages 18-40 (mean 22.3), volunteered. All participants were right handed and half had no experience with in air interactions.

Task & Design – The experimental task was the same as in experiment one. The independent variables were *plane size*, and *selection method*. We used a 2×4 within-subject design. We tested two plane sizes: 35x35 cm and 25x25 cm. Selection method indicates the mechanism by which participants select a target. We designed four methods:

- *Click*: as in experiment one; participants click a mouse held in their left hand.
- *Swipe*: is a quick horizontal arm movement to both sides at min 50 cm/sec and for a movement of at least 15 cm.
- *Dwell*: participants highlight a target for 1.5 seconds (threshold determined through iterative pilot testing).
- *Second Hand*: participants move the left arm 20cms away from its resting position (i.e., from the hips).

The experimenter demonstrated each selection method and participants had an initial training with each condition, testing each selection method until they had control over it. The experiment had a total of 2×4 = 8 conditions and each condition had 3 trials, yielding 2×4×3 = 24 trials per participant, or 288 trials in total. All participants completed the experiment within approximately 45 minutes. The trials were counter-balanced with a Latin-square approach on selection method and plane size appeared in a random order.

Measures –We collected values for CE, completion time, and landing error rate. Participants filled in a Borg CR10 scale after each condition.

Results

We used the same statistical tests as in experiment one (ART ANOVA). Figure 6 shows an overview of the results.

Consumed Endurance (CE) – Results showed a main effect of *selection method* ($F_{3,33}$ = 19.612, $p < 0.001$) and *plane size* ($F_{1,11}$ = 16.165, p = 0.002). There were no significant interaction effects for *plane size* × *selection method* (p = 0.323). Post-hoc pair-wise comparisons on *selection method* yielded significant ($p < 0.04$) differences between all pairs except between dwell and swipe, and dwell and second hand. In general, CE was lowest for click at 8.18% (std = 9.06) and the small plane at 10.80% (std = 12.00).

Borg CR10 – Results showed a main effect of *selection method* ($F_{3,33}$ = 8.425, $p < 0.001$) but not for *plane size* (p = 0.837). Results did not show interaction effects for *plane size* × *selection method* (p = 0.586). Post-hoc analysis on *selection method* yielded significant differences for all pairs ($p < 0.04$) except second hand and dwell. In general, Borg CR10 was lowest for click at 1.188 (std = 0.845).

Completion Time – Results showed main effects of *selection method* ($F_{3,33}$ = 58.076, $p < 0.001$) and *plane size* ($F_{1,11}$ = 5.143, p = 0.044). Analysis also revealed interactions effects for *plane size* × *selection method* ($F_{3,33}$ = 3.167, p = 0.037). Post-hoc analysis on *selection method* revealed significant differences ($p < 0.017$) between all pairs. Click was the fastest selection method at 40.76 seconds (std = 10.62). The larger plane had a lower mean completion time of 54.52 seconds (std = 17.92).

Landing Error Rate – Results showed a main effect of *plane size* and *selection method* (all $p<0.003$) on error rate with $F_{1,11}$ = 17.112 and $F_{3,33}$ = 22.167 respectively. There were no interaction effects between *plane size* and *selection method* (p = 0.131). Post-hoc pair-wise analysis showed a significant difference between swipe and all other selection methods (all $p<0.001$). Error rate was lowest for dwell at 0.66 (std = 0.59) and the big plane at 0.83 (std = 0.86).

Discussion

This experiment highlights the capacity of CE for uncovering differences where subjective ratings cannot. A larger plane requires stretching and lifting the arm which clearly results in increased effort. CE reveals a significant difference between plane sizes which Borg CR10 hides due to the high variance and small size of the sample.

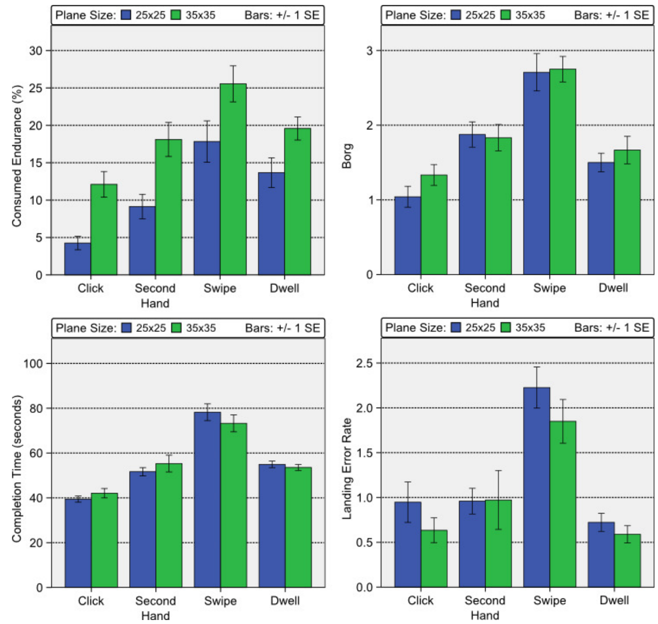

Figure 6. Consumed Endurance, Borg CR10, completion time and error rate for experiment two.

Selection methods which do not require movement of the selecting hand perform best across all metrics. Swipe, which performs worst, sees its CE increased due to the greater amount of movement it requires due to the gesture design and to tracking errors. Tracking errors, more noticeable in the small plane, are due to problems of distinguishing the arm from the body and to follow the hand back (such as in swipe). This results in poorly controlled gestures which miss the target, leading to repetition, and therefore higher completion time and CE. A better tracking technology would increase the controllability of the gesture, reducing the need to correct and flattening error rates and their effect on CE.

The best plane in terms of CE is the small plane. However, the best performance in terms of completion time and landing error rate is the big plane. A designer may have to choose the larger plane to reduce errors which could quickly lead to fatigue and a bad user experience. As expected, Click outperforms all other selection methods in terms of CE and therefore it should be used when possible, else Dwell and Second Hand use similarly little CE.

ENDURANCE AS A DESIGN PARAMETER

The previous experiments demonstrate the use of CE as a tool to assess various design alternatives. In this section we use another endurance-related metric, *strength* (defined earlier), as a design parameter for a mid-air text-entry system. We choose text-entry because it is a common task and one that involves repetition. From our previous experiments we know that: (a) interactions consumed the least endurance when they occur on the center plane with a bent arm; (b) a 25x25 cm plane size consumes lower CE; and (c) for single hand situations dwell selections are recommended.

In this section we propose a new text-entry layout optimized for such a set of interaction parameters (see Figure 7). We collected data from 4 participants (all male) who held the cursor at each position of the 6x6 grid for 10 seconds (center

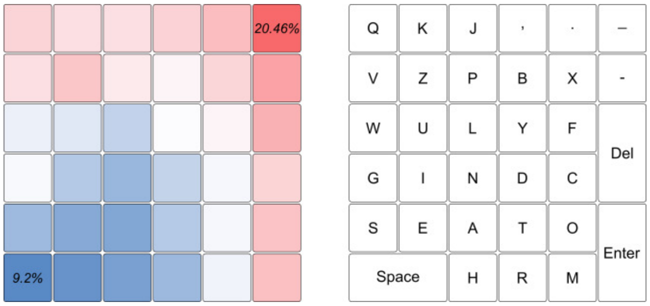

Figure 7. Left: Heat-map of strength. Cells in blue require the least strength and those in red require the most. Right: SEATO key layout based on character probability in the English language and strength.

plane, arm bent, 25x25 cm plane size). We recorded arm *strength* ($\frac{\|\vec{T}_{shoulder}\|}{T_{max}} * 100$) for each cell because, unlike endurance and CE, it is not affected by the asymptote of equation 2 and thus reveals differences even when the physical effort is low. Figure 7-left shows the resulting heat-map for strength throughout the grid: on average. The cell on the lower-left corner requires 9.2% of the maximum strength, while the cell on the upper-right corner required 20.46%. All bluish cells in Figure 7-left are below the 15% threshold. Figure 7-right shows the resulting SEATO text-entry layout for mid-air interactions. We obtained the SEATO layout by mapping the cells with the lowest strength demands to the characters with the highest probability in the English language, ideally resulting in a less physically demanding interaction than with other text-entry layouts like QWERTY.

Experiment 3 - Text Entry Layout

In this experiment we compare the SEATO and QWERTY layouts in terms of CE, text-entry speed and error rate.

Methods

Subjects – 12 participants (5 female), ages 18-40 (mean 24), volunteered. All participants were right handed and all but three had previous experience with mid-air interactions.

Task – Participants had to type a sentence that was shown on the screen. For typing a character participants had to move the cursor to the cell with the character and use the dwell gesture for selection. We selected a list of 53 sentences between 19 and 23 characters long from MacKenzie et al.'s set [24]. When the wrong character was selected the system would not allow any more typing until the wrong character is deleted by selecting the DEL key; this is counted as an entry error. The task finishes when the correct phrase is typed in and the participant selects the ENTER key.

Design – The independent variable is layout: SEATO and QWERTY. We used a within-subjects design to compare CE between layouts. Participants had an initial training with the SEATO layout and with the mechanics of selecting a letter. Participants were trained by typing sample sentences with both layouts, terminating a phrase with the ENTER key. There were 2 conditions, and each condition had a total of 4 blocks and 3 trials per block, yielding 2×4×3 = 24 trials per participant, or 288 trials in total.

Measures – We measured CE, words per minute (WPM), and error rate. Users filled a Borg CR10 scale after each block.

Results

We used the same analysis tools as for experiments one and two. Figure 8 shows an overview of the results.

Consumed Endurance (CE) – Results showed a main effect for *layout* ($F_{1,11} = 51.332$, $p < 0.001$) and *block* ($F_{3,33} = 14.285$, $p < 0.001$). Results did not show significant interaction effects ($p = 0.174$). Post-hoc analysis showed a significant difference between the first and second block ($p = 0.003$). SEATO had a lower average CE compared to QWERTY at 6.43% (std = 11.08).

Borg CR10 – Results did not show a main effect of *layout* ($p = 0.258$) or *block* ($p = 0.257$). Interactions effects were also not significant ($p = 0.300$).

Words Per Minute – Results showed no main effect of *layout* on words-per-minute (WPM) ($p = 0.124$), but a main effect for *block* on WPM ($F_{3,33} = 6.120$, $p=0.002$). Results showed no *layout* × *block* interaction effect ($p = 0.581$). Post-hoc analysis revealed significant differences between the first and third ($p < 0.02$) and last ($p = 0.003$) blocks. The last block had the highest WPM at 4.55 (std = 1.26).

Typing Error – Results showed a main effect of *layout* ($F_{1,11} = 15.868$, $p = 0.002$) and *block* ($F_{3,33} = 8.572$, $p < 0.001$), but no interaction effects ($p = 0.378$). Post-hoc analysis revealed the first block to be significantly different from all other blocks ($p < 0.022$). The last block had the lowest mean error rate at 0.05 (sd = 0.07) and the Qwerty layout had a lower mean error rate at 0.06 (sd = 0.1).

Discussion

Our data shows that layout has an effect on CE, with our proposed SEATO layout consuming significantly less endurance than QWERTY (a quarter), at no cost in terms of words-per-minute and only slightly higher error rate. Moreover, results show no significant difference in the Borg CR10 rankings, outlining the added value of our metric for situations where differences do not surface with subjective ratings. Finally, the similar typing speed we observed reinforces the notion that designers could also look at other factors beyond interaction time for making interface choices.

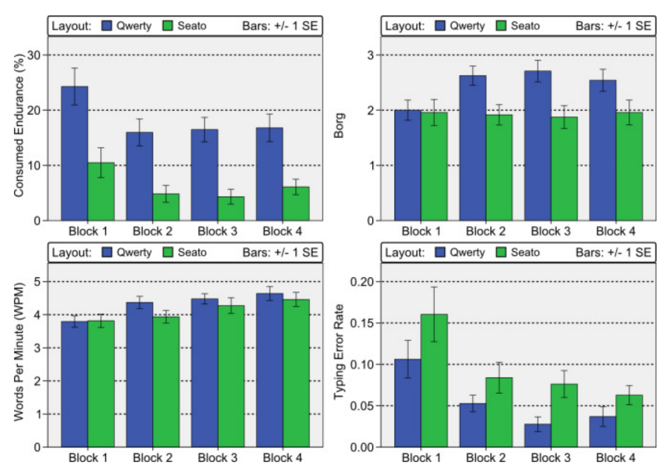

Figure 8. Consumed Endurance, Borg CR10, words per minute and error rate for experiment three.

GENERAL DISCUSSION

We discuss our findings in light of mid-air interactions.

Applications of endurance-based metrics

Our results demonstrate the value of adopting CE as a complementary guide for evaluating the impact of mid-air input parameters like plane size or selection mechanisms on fatigue. In a similar vein, CE can be used to evaluate alternative sets of mid-air gestures for controlling an interactive system (as in Barclay et al. [6]).

Aside from designing endurance-efficient text-entry layouts, our metrics can be used in the design of mid-air menus, document navigation controls and arm gestures. Based on the heat-map shown in Figure 7-left, when selecting a menu with a pointer, the most frequently used menu items should be in the lower left corner (or lower right if interacting with the left arm): buttons on the top or the right side of the interaction plane should be avoided (Figure 9-middle). Similarly, navigation controls, if used frequently, should appear in those regions marked in blue in Figure 7 (see image in color).

Our results also suggest that when possible mid-air gestural interactions should consider relative movements rather than absolute ones that have fixed positions in the air (Figure 9-right). In this manner, gestures could take place in regions of least effort. For example, gesturing the letter 'B' could take place by allowing users to start the gesture by moving the arm from its rest position without having to lift it up to an absolute start position of engagement.

Finally, to control for arm position (bent or extended) in our experiments, our application did not allow the user to operate outside a certain distance region. While we do not advise enforcing such restrictions in mid-air interactions, application designers could include guidelines to users, in the form of a quick image or video clip, to reduce fatigue during use.

While we demonstrated the use of our metric to minimize CE, other applications may choose to increase it or adjust it dynamically. For example, mid-air gaming applications could introduce CE for better control over game balancing. Dynamic game-balancing is possible by gradually shifting the need for selecting or interacting with different positions within the interaction plane or by requiring the user to use different arm positions (switching between extended and bent). This could have direct benefits in virtual therapy applications where movements can become increasingly demanding as the patient's upper limb functions improve, or conversely if the patient's progress is slow.

To support the different explorations and usages of CE, researchers and designers of mid-air interfaces can download our implementation here http://hci.cs.umanitoba.ca/projects-and-research/details/ce.

Our findings complement existing guidelines

Our results, obtained with a view on reducing fatigue, empirically confirm and further complement human interface guidelines proposed by some manufacturers of gestural tracking systems (for details see: www.microsoft.com/en-us/kinectforwindows/develop/learn.aspx). Such guidelines mainly provide designers with parameters for optimal tracking efficiency. For example, the Kinect guidelines suggest using Dwell to avoid inadvertent selections (page 55, in above document) and recommend that gestural systems allow seamless hand switching or provide alternative gesture sets to reduce fatigue (page 22). Our results further provide specific insight on how such alternative gesture sets should be designed to reduce effort, such as for text-entry.

Our findings in light of previous results

Our results justify the fatigue-related findings of prior work. Harrison et al.'s participants preferred a position with "elbows tucked in, hands held front, and palms up" [20]. In light of our results that position seems natural as it closely resembles the center bent arm position. Similarly Boring et al.'s participants who did not move their whole arms and relied on tilt reported less fatigue [9]. This result also seems natural as the arm was not fully extended and thus all of its mass did not have to be moved by the shoulder muscles. Finally, our results can explain why Cockburn et al.'s ray-casting technique was ranked the least physically demanding [11], as the upper-arm was held in a resting position.

Our results can also be used to re-consider existing interactions. For example, Li et al.'s VirtualShelves introduce mobile interactions across the horizontal and vertical axis in front of the user [23]. As these movements require full arm extension their CE is high. An endurance-efficient alternative can use only movements of the forearm, with the upper arm in rest position, i.e. with bent elbow. Similarly, Cockburn et al.'s 2D plane technique can improve in terms of CE by fixing their interaction plane at the center plane location, i.e. between the hip and shoulders, and by reducing the size of the plane to one where less arm extension is needed.

Lessons learned

We take away these lessons from our initial exploration:

- The center + bent arm position for selections on a 2D plane is the least tiring of all positions we tested.
- The regions at the bottom of the interaction plane improve CE. Interacting in the lowest possible region should be dictated by the tracking system's accuracy.
- In the center bent arm position a bigger plane can be used to reduce tracking-induced errors.
- A clicking device for selection minimizes fatigue. When only one arm is available, the *dwell* method is best.

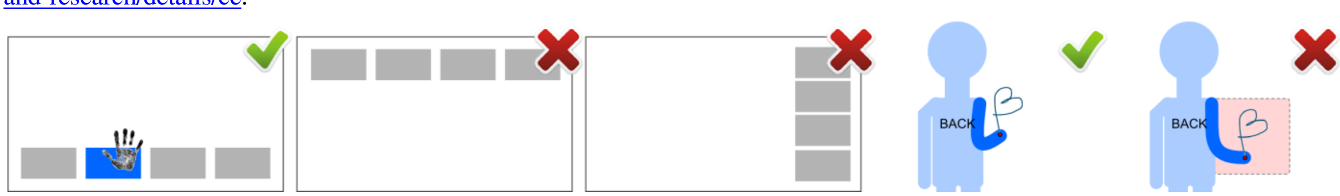

Figure 9. Design implications of consumed endurance. Left, menu items should be located in the bottom of the UI. Right, for some applications, such as free gesturing, designers may consider relative input which is possible anywhere in the interaction plane instead of a fixed location.

- Time-based metrics are incomplete indicators of fatigue.
- Strength can be used to inform the design of endurance-efficient interactions techniques.
- The SEATO layout supports endurance-efficient mid-air text-entry, without compromising efficiency.

LIMITATIONS

Our CE implementation for Microsoft Kinect presents two main limitations. First, it requires line of sight to the user's complete body in order to form a complete skeleton. Second, the skeleton measurements become noisy due to difficulties differentiating between the user's arm and body (especially when the arm is close to the body). These difficulties can be avoided in future versions of the sensors (higher resolution, improved tracking) or using alternative tracking systems.

In future work, we will extend our model to capture other arm-segments and use individual body metrics (length and mass). Moreover, while this paper shows a strong correlation between CE and Borg CR10 during simple mid-air arm movements; further research is needed into highly dynamic settings and the effects of experience and accumulated fatigue. Finally, as advances in the fields of sport sciences and ergonomics refine the notion of muscle fatigue in an objective manner, the definition and validity of CE should also be revisited against such objective metric.

CONCLUSIONS

In this paper we introduce *consumed endurance*; a metric to characterize shoulder fatigue in mid-air interactions. CE only requires the tracking system used to interact with the NUI itself, and thus it is a real-time, objective, non-invasive and non-obtrusive approach to assess gorilla-arm. Through an initial study, we showed CE's validity as a metric of fatigue and its gender neutrality. Using CE, researchers do not need to ask participants about their perceived physical effort due to the strong correlation between CE and the Borg CR10 scale.

We showed how CE can be used as an evaluation tool for selecting suitable mid-air interaction parameters. We focused our exploration on item selection in a 2D plane and investigated the suitable variables for *plane location*, *arm extension*, *plane size* and *selection method*. Our results show that the combination of plane location and arm extension with the least endurance demands (i.e., creating the least fatigue) is at the vertical center of the body, on the side of the moving arm, and with a bent posture. Finally, selections by the dwell method are most appropriate when only one hand is available. Our results along with a related metric, *strength*, guided the design of the SEATO text-entry layout for mid-air interactions. Results show that SEATO is on par with QWERTY in terms of words per minute and typing error rate, and consumes only a quarter of endurance.

APPENDIX A – FORCE AND TORQUE CALCULATIONS

From the definition of torque ($\vec{T}$) we know that $\vec{T} = \vec{F}d\sin\theta$. Where d is the distance from the shoulder joint to where the force is applied, θ is the angle between the force vector and the axis, and F is the measured force at distance d. Given that all forces are tangential to the distance vector, we know that $\sin\theta = 1$. Therefore, the equivalent F_{max} at the CoM at distance r ($F_{max,r}$) is:

$$F_{max}d_{elbow} = F_{max,r}r \quad \rightarrow \quad F_{max,r} = \frac{F_{max}d_{elbow}}{r} \qquad (8)$$

The $\frac{F}{F_{max}}$ ratio at the CoM can be expressed as:

$$\frac{\|\vec{F_r}\|}{F_{max,r}} = \frac{\|\vec{F_r}\|}{\frac{F_{max}d_{elbow}}{r}} = \frac{\|\vec{F_r}\|r}{F_{max}d_{elbow}} = \frac{\|\vec{T}\|}{T_{max}} \qquad (9)$$

APPENDIX B - ARM CENTER OF MASS CALCULATION

The CoM of a two segment body is located along the vector linking the CoMs of each segment, at a distance from the first segment's CoM equal to the ratio between the second segment's mass and the combined masses of both segments. Figure 10 shows the arm as a three segments body composed of upper arm (ShEb), forearm (EbWr), and hand (WrHa). Applying the process described above for a two segment body, and using the values presented by Freivalds [17], we calculate the CoM of the forearm + hand combination as:

$$D = B + \frac{WrHa_{mass}}{EbWr_{mass} + WrHa_{mass}}\overrightarrow{BC} \qquad (10)$$

Then, we apply a similar process for the CoM of the upper arm + (forearm + hand) combination as:

$$CoM = A + \frac{EbWr_{mass} + WrHa_{mass}}{Arm_{mass}}\overrightarrow{AD} \qquad (11)$$

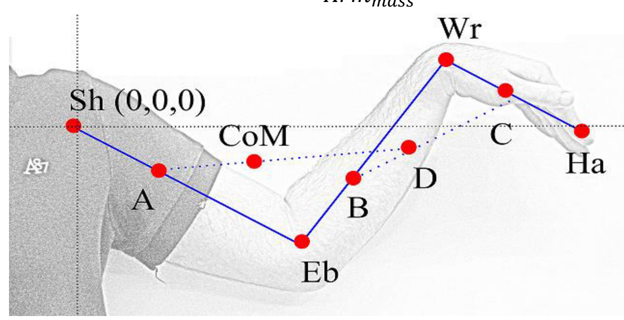

Figure 10. Arm segments involved in calculating its CoM.

APPENDIX C - ARM INERTIA CALCULATION

The inertia of a multi-segment body like the arm ($\overrightarrow{I_{arm}}$) is a vector of magnitude equal to the sum of each segment's inertia, and in the direction (unit vector = $U_{\overrightarrow{I_{arm}}}$) of the cross product of the movement of its CoM:

$$\overrightarrow{I_{arm}} = U_{\overrightarrow{I_{arm}}} * \|\overrightarrow{I_{arm}}\| \qquad (12)$$

Where the unit vector of the direction ($U_{\overrightarrow{I_{arm}}}$) is equal to:

$$U_{\overrightarrow{I_{arm}}} = \frac{\overrightarrow{ShCoM_{prev}} \times \overrightarrow{CoM_{prev}CoM}}{\|\overrightarrow{ShCoM_{prev}} \times \overrightarrow{CoM_{prev}CoM}\|} \qquad (13)$$

And the magnitude ($\|\overrightarrow{I_{arm}}\|$) is equal to:

$$\|\overrightarrow{I_{arm}}\| = \|\overrightarrow{I_{upper}}\| + \|\overrightarrow{I_{fore}}\| + \|\overrightarrow{I_{hand}}\| = 0.0201 \qquad (14)$$

REFERENCES

1. Gesture recognition. http://en.wikipedia.org/wiki/Gesture_recognition.
2. Gorilla arm. www.computer-dictionaryonline.org/index.asp?q=gorilla+arm.
3. Visible human project. www.nlm.nih.gov/research/visible/visible human.html.
4. Bachynskyi, M., Oulasvirta, A., Palmas, G., and Weinkauf, T. 2013. Biomechanical simulation in the analysis of aimed movements. In CHI '13 EA. ACM.
5. Bakke, M., Thomsen, C., Vilmann, A., Soneda, K., Farella, M., and Miller, E. 1996. Ultra sonographic assessment of the swelling of the human masseter muscle after static and dynamic activity. Archives of Oral Biology 41, 2, 133 – 140.
6. Barclay, K., Wei, D., Lutteroth, C., and Sheehan, R. 2011. A quantitative quality model for gesture based user interfaces. In Proc. OZCHI '11, ACM, 31–39.
7. Bigdelou, A., Schwarz, L., and Navab, N. 2012. An adaptive solution for intra-operative gesture-based human-machine interaction. In Proc. IUI '12, ACM.
8. Borg, G. 1998. Borg's Perceived Exertion and Pain Scales. Human Kinetics.
9. Boring, S., Jurmu, M., and Butz, A. 2009. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In Proc. OZCHI '09. ACM.
10. Bustamante, E.A., and Spain, R.D. 2008. Measurement invariance of the Nasa TLX. In Proc. of the Human Factors and Ergonomics Society Annual Meeting 52, 19.
11. Cockburn, A., Quinn, P., Gutwin, C., Ramos, G., and Looser, J. 2011. Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. Int. J. Hum.-Comput. Stud. 69, 6, 401–414.
12. Cuccurullo, S., Francese, R., Murad, S., Passero, I., and Tucci, M. 2012. A gestural approach to presentation exploiting motion capture metaphors. In Proc. AVI '12, ACM.
13. Dedering, Å., Gnospelius, Å., and Elfving, B. 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), 189-198.
14. Edmunds, T., Gentner, R., d'Avella, A., and Pai, D. 2012. Feasible wrench space and its estimation for isometric haptic interaction. In Proc. HAPTICS '12. IEEE.
15. Elfving, B. and Dedering, Å. 2007. Task dependency in back muscle fatigue – Correlations between two test methods. Clinical Biomechanics, 22(1), 28-33.
16. Ferguson, S.A., W. Gary, A., Le, P., Rose, J.D. and Marras, W.S. 2011. Shoulder Muscle Oxygenation During Repetitive Tasks. Human Factors and Ergonomics Society Annual Meeting 55,1, 1039-1041.
17. Freivalds, A. 2004. Biomechanics of the Upper Limbs: Mechanics, Modeling, and Musculoskeletal Injuries. 1 ED. CRC Press.
18. Garner, B.A., and Pandy, M.G. 2001. Musculoskeletal model of the upper limb based on the visible human male dataset. Computer Methods in Biomechanics and Biomedical Engineering 4, 2, 93–126.
19. Garzotto, F., and Valoriani, M. 2012. Don't touch the oven: motion-based touchless interaction with household appliances. Proc. AVI'12, ACM.
20. Harrison, C., Ramamurthy, S., and Hudson, S.E. 2012. On-body interaction: armed and dangerous. In Proc. TEI'12, ACM.
21. Hart, S.G., and Stavenland, L.E. 1998. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human Mental Workload, P. A. Hancock and N. Meshkati, Eds. 139-183.
22. Huang, M.-C., Chen, E., Xu, W., and Sarrafzadeh, M. 2011. Gaming for upper extremities rehabilitation. In Proc. WH '11, ACM.
23. Li, F.C.Y., Dearman, D., and Truong, K. N. 2009. Virtual shelves: interactions with orientation aware devices. In Proc. UIST '09, ACM.
24. MacKenzie, I.S., and Soukoreff, R.W. 2003. Phrase sets for evaluating text entry techniques. In CHI EA '03, ACM.
25. Mentis, H. M., O'Hara, K., Sellen, A., and Trivedi, R. 2012. Interaction proxemics and image use in neurosurgery. In Proc. CHI '12, ACM.
26. Morris, M. G., Dawes, H., Howells, K., Scott, O. M., and Cramp, M. 2008. Relationships between muscle fatigue characteristics and markers of endurance performance. Sports Science and Medicine 7, 431-436.
27. Öberg, T., Sandsjö, L., and Kadefords, R. 1994. Subjective and objective evaluation of shoulder muscle fatigue. Ergonomics, 37(8), 1323-1333.
28. Peres, S.C., Nguyen, V., Kortum, P.T., Akladios, M., Wood, S.B., and Muddimer, A. 2009. Software ergonomics: relating subjective and objective measures. In CHI EA '09, ACM.
29. Rohmert, W. 1960. Ermittlung von erholungspausen fr statische arbeit des menschen. European Journal of Applied Physiology and Occupational Physiology 18. 123–164.
30. Sjgaard, G., Savard, G., and Juel, C. 1988. Muscle blood flow during isometric activity and its relation to muscle fatigue. European Journal of Applied Physiology and Occupational Physiology 57, 327–335.
31. Tan, H., Radcliffe, J., Ga, B.N., Tan, H.Z., Eberman, B., Srinivasan, M.A., and Cheng, B. 1994. Human factors for the design of force-reflecting haptic interfaces. Dynamic Systems and Control, 55(1), 353-359.
32. Troiano, A., Naddeo, F., Sosso, E., Camarota, G., Merletti, R., and Mesin, L. 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), 179-186.
33. Wobbrock, J.O., Findlater, L., Gergle, D., and Higgins, J.J. 2011. The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. In Proc. CHI '11. ACM, 143-146.