Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions

Jeffrey T. Hansberger[1(✉)], Chao Peng[2], Shannon L. Mathis[2],
Vaidyanath Areyur Shanthakumar[2], Sarah C. Meacham[2],
Lizhou Cao[2], and Victoria R. Blakely[2]

[1] Army Research Laboratory, Huntsville, AL, USA
`jeffrey.t.hansberger.civ@mail.mil`
[2] University of Alabama Huntsville, Huntsville, AL, USA

Abstract. The use of gestures as a way to interact with computer systems has shown promise as a natural way to interact and manipulate digital information. However, users performing mid-air gestures for even moderate periods of time experience arm fatigue and discomfort, earning its name of the gorilla arm syndrome. Based on the natural use of hands during communication, a new gesture vocabulary was created that supports the arms while the user performs the gestures. A repeated measures within subject design was conducted where participants interacted with a custom video game using 3 types of input for 30 min each, (1) keyboard, (2) mid-air gestures and (3) supported gestures. Three measures of exertion were collected, (1) time, (2) energy expenditure, and (3) perceived exertion. The newly designed supported gestures required significantly less physical and perceived effort than the mid-air gestures and required similar exertion as the keyboard condition.

Keywords: Gorilla arm · Mid-air gestures · Fatigue · Endurance · Gesture interaction · Natural user interface

1 Introduction

The idea of using hand and arm gestures to interact with computer systems is not a new idea. The "Put-That-There" study in 1980 [5] defined and used hand gestures for a graphical user interface. Hollywood further popularized the idea in 2002 with the gesture based interfaces that Tom Cruise interacted with in the "Minority Report" movie and more recently, the Tony Stark character in the "Iron Man" movie series. Even though the "Minority Report" interface was based on research by Underkoffler [30] and has captured the imaginations of both the public and the HCI community, the use of gestures as the primary mode of interaction for an interface remain more of a novelty (e.g., Microsoft Kinect games). One of the primary reasons for this is the issue commonly called the gorilla arm syndrome [7].

The gorilla arm syndrome originally arose with the advent of touchscreens used in a vertical orientation, which forces the user to extend their arms without support. When this is done for any task longer than a few minutes (e.g., with an ATM), it causes arm fatigue and a feeling of heaviness in the arms. Mid-air gestures, just as the name suggests, suffers from the same issue of an unsupported arm position. The common position for both vertical touchscreen use and mid-air gesture input is with the arm/s extended in front of the user around shoulder height (Fig. 1A). Hincapie-Ramos, *et al.* [20] found this position to be the most fatiguing physical position out of the 4 they investigated. One reason for the necessity of this arm position is due to the technology used to detect the user's arm and hand position.
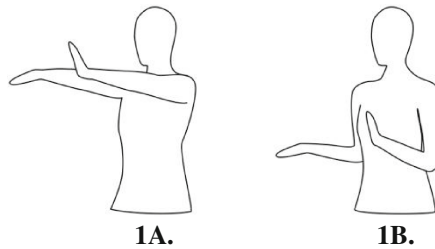
Fig. 1.   Posture 1A. is a common arm position with high fatigue for vertical touchscreen and mid-air gestures. Posture 1B. is the arm position with low fatigue for gestures during face-to-face communications.

The challenge of capturing the user's arm and hand position has been met with a number of technology approaches. Optical solutions with external cameras that track the user's motion can include two basic types, a marker based system and markerless motion capture. The marker based system uses input from multiple cameras to triangulate the 3D position of the user wearing special markers while the markerless motion capture uses one or more cameras and computer vision algorithms to identify the user's 3D position. For issues of practicality, the markerless motion capture represents the optical motion capture of choice for general use. The Microsoft Kinect and the Leap Motion sensor are examples of markerless motion capture systems that are both affordable and accessible to consumers, researchers, and developers. However, these type of optical sensors must have an unobscured view of the user's hands, which often forces the user's hands into that high fatigue area in front of the user mentioned above and away from many natural arm supports like the arm rest of a chair. In addition, these sensors have shown to be limited in their gesture recognition accuracy and reliability (e.g., [8, 18]).

Another approach that does not use any optical devices like cameras is a system that uses inertial measurement units (IMUs). The IMU approach consists of several sensors placed on the user or in the clothing the user wears. Each IMU consists of a gyroscope, magnetometer, and accelerometer in order to wirelessly transmit the motion data of the user to a computer, where it is translated to a biomechanical model of the user. IMU gloves and suits have typically been used by the movie and special effects industry but recent crowdsourcing efforts like the Perception Neuron IMU

suit have provided more affordable IMU based motion capture solutions. IMU solutions do require the user to wear the sensors in the form of gloves or straps but unlike the optical solutions, it does not provide constraints on where the user's hands must be to perform the gestures. As long as the sensors are within Wi-Fi range of the router, there are no constraints on the position, orientation, or worry of obscuring the hands from an external camera source.

Without these line-of-sight constraints that are required by optical motion capture systems, we have created a new approach that is based on the natural and non-verbal ways people use their hands to communicate to one another. The use of IMU technology for motion capture has allowed us to create gestures that mimic how people naturally use their hands, which avoids the highly fatiguing arm positions that cause gorilla arms. This paper will describe these new gestures, an experiment comparing the fatigue levels across keyboard, mid-air, and our newly formed hand gestures, and discuss the implications of this work.

2   Background

The name of mid-air gestures itself alludes to the fatigue problem associated with these types of gestures. Suspending the arms in mid-air without support will always pose a physical challenge for the user over prolonged periods of interaction. Fatigue levels are determined as the amount of time a person can maintain an isometric (static) muscle contraction [9]. This can be measured by the heart's response to increase blood flow to transport oxygen to the muscle fibers (cells). An isometric muscle contraction causes an impairment of blood flow due to the increase in intramuscular pressure. Complete vascular occlusion occurs at 70% of a person's maximal voluntary contraction (MVC). At approximately 50% of MVC, a user has approximately 1 min until fatigue. Muscle fatigue is caused by a reduction in the amount of oxygen transported to the muscle fibers. Oxygen is needed for aerobic energy metabolism which can be maintained for hours. When there is not enough oxygen to support the energy needs of the muscle, energy is provided by partial or complete conversion to the glycolysis process (anaerobic metabolism), in which the muscle fiber consumes glucose for energy. As the muscle cells consume the limited stored glucose to maintain the current position, the user begins to feel the effects of muscle fatigue. This fatigue increases until the glucose is consumed and the user can no longer maintain the arm position. At the same time, an increase in heart rate and blood pressure is needed to maintain blood flow. As a result, oxygen consumption (VO_2), an estimate of a person's energy requirement to perform muscular work, will increase above the resting value. At rest, a 70 kg person will have a VO_2 of approximately 3.5 ml of O_2 per kilogram of body weight per minute ($mL \bullet kg^{-1} \bullet min^{-1}$). As the level of fatigue increases, VO_2 will also increase.

As mentioned before, the arm position that produces a high level of fatigue is often the one required by many gesture based systems (Fig. 1A) [20], particularly with one of the most common gestures of pointing [8]. The least fatiguing arm position is shown in Fig. 1B, which has the user's arm bent at the elbow with their hands extended to the front. This is also the same basic position of the arms and hands when people use hand

gestures to complement their speech during communication exchanges [21]. The use of hand gestures during speech is so natural and ubiquitous that people gesticulate as much whether the person they are talking to can see them or not [10].

Hand gestures by the speaker improves their ability to learn new concepts or improve their ability to teach a concept to someone else [26]. It is therefore not a surprise that the speakers of some of the most widely viewed TED talks use more than 600 hand gestures in a single 18-minute presentation [14]. Professional speakers and professors that must speak to audiences for 1–3 h at a time are performing large quantities of unsupported hand gestures. How are they doing this without extreme fatigue in their arms and shoulders after performing gestures for hours? The answer is that these speakers are conducting most of their hand gestures in the comfortable zone shown in Fig. 1B. They are also periodically relaxing their arms by extending them straight down, in a pocket, or taking advantage of a supporting structure like a podium. This periodic relaxation of the arms allows improved blood flow and therefore oxygen to the previously contracted muscles, which avoids switching to a glycolysis process that causes extreme arm fatigue (i.e., gorilla arms).

We have adopted this natural use of hand gestures to construct a limited gesture vocabulary for users. Our specific use case is a knowledge worker sitting at a desk and interacting with large amounts of visual images but could apply to other domains. We are applying 2 strategies to avoid the fatigue brought about from the gorilla arm syndrome by (1) placing the arms in the least fatiguing position with them bent at the elbow (Fig. 1B.) and (2) using support in the form of the user's lap or on the arm rest of the chair (Fig. 2). Other researchers have proposed resting the user's elbows on a surface (e.g., [8, 15, 17, 28]) but none have conducted a systematic examination on the physical and perceived fatigue caused by mid-air and supported gestures over even moderate durations of time.
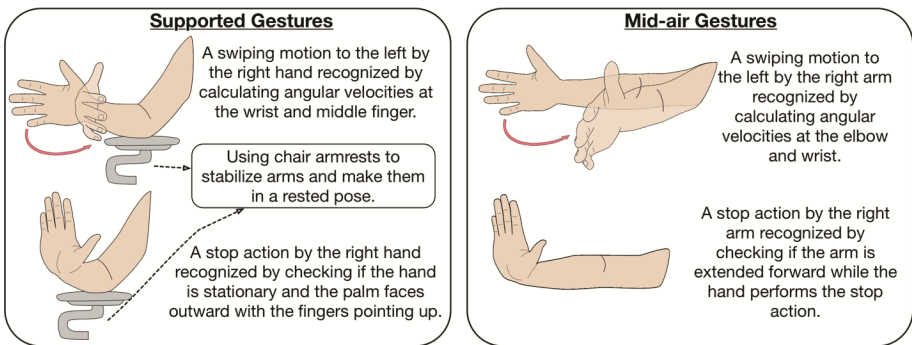
Fig. 2. Supported and mid-air gestures.

Most prior research also uses a table top to support the user's arms (e.g., [8, 17]). Similar to the ergonomic recommendations for proper keyboard positioning that has found for every 1 cm increase in keyboard height, neck and shoulder discomfort increases 18% [4, 22], we have avoided making the user place their arms at or above table level. Our supported gesture position would be difficult for most optical motion

capture systems due to the handsplaced below the surface of the table but is easily supported with wireless IMU glove technology. We have therefore selected the technology that will support our use case instead of allowing the technology to dictate the actions and postures of our users.

2.1 Supported Gesture Vocabulary

We scoped the gesture requirements to encompass 2 basic usability engineering principles [24, 25]: (1) the gestures should minimize user cognitive load of memorizing how to perform them and (2) the gestures should clearly correspond to the user's natural behaviors. The use case involves a knowledge worker manipulating digital images within a large image collection. We have designed a supported gesture vocabulary that contains 10 gestures for the final prototype. As discussed below, the gestures are general enough to apply to a number of different tasks or domains. To examine the viability of prolonged use of these gestures before fully developing the prototype, we have developed a game that uses 3 of the gestures. The gestures include (1) swipe left, (2) swipe right, and (3) stop.

Rather than relying on machine learning algorithms to detect gestures, we propose a novel gesture recognition framework that relies on the vectors between 2 sensors. Researchers have obtained good results in mid-air gesture recognition using machine learning algorithms. However, this requires a large amount of training data and manual annotation of the data as ground truth in order to train accurate gesture recognition models [2, 23]. This type of approach usually considers static hand gestures but not the gestures involving hand movements (e.g., [19, 29]). Our approach computes angular velocities between vectors to recognize both static and dynamic gestures and avoids some of the drawbacks of the machine learning algorithm approach.

The typical motion capture skeletal hand model consists of 16 joints in each hand, 3 on each finger and 1 at the wrist. Based on the motion of these joints during natural swipe and stop hand gestures, we used the positions of 4 hand joints and 2 arm joints. Vectors are formed using the positions of joints on the hands and arms. Among the joints with sensors, we identified 3 relevant vectors in each hand and 2 relevant vectors in each arm (Fig. 3). By monitoring the angular velocity of the pre-identified vectors, we can accurately recognize the following gestures.
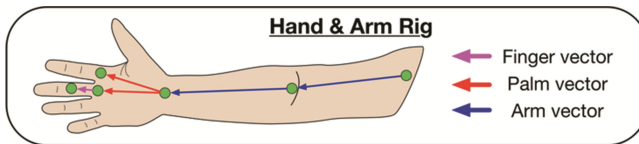
Fig. 3. IMU sensors are placed at the joints of the hand and arm to capture position values in real-time.

Supported Gesture 1 & 2 – Swipe Left & Right. A supported swiping motion of the hand to the left by the right hand or to the right with the left hand indicates movement of the target in that direction. This motion is primarily accomplished at the wrist so the

arms can easily be supported by the arm rest of a chair or the participant's lap. We define the start state of a swipe gesture if the palms of each hand are facing each other. Then if the angular velocity of the vector from the wrist to the middle finger exceeds our defined threshold, it is considered a swipe gesture in that direction.

Supported Gesture 3 – Stop. A supported stop gesture provides a stop command to the target object. This common gesture is characterized by facing the palm of the hand outward with the fingers pointing up. Again, this gesture is easily accomplished while the arms are at rest. Recognition of the stop gesture uses the angles between the finger and wrist sensors to activate the command. For this experiment, the target continues its current motion or trajectory until the stop gesture is removed.

2.2   Mid-Air Gesture Vocabulary

Mid-air gestures typically involve the use of the entire arm for the reasons indicated earlier. The same 3 gestures were created as mid-air gestures using the vector velocity and angle approach to conduct (1) swipe left, (2) swipe right, and (3) stop actions.

Mid-Air Gesture 1 & 2 – Swipe Left & Right. Amid-air swiping gesture to the left by the right hand and to the right with the left hand was created. This gesture uses the shoulder, arm, and hand vector to define the starting position with the arms extended in front of the participant at shoulder level. The vector velocity of the hand and arm is then monitored to detect the swiping motion done by bending the arm at the elbow to the left or right.

Mid-Air Gesture 3 – Stop. Amid-air stop gesture is defined in the same manner as the supported stop gesture but with the arm extended in front of the user.

2.3   Objective

The objective of this study is two-fold. The first objective is to create hand gestures that leverage how we naturally position, relax, and support our hands and arms to reduce the level of exertion and fatigue. The second objective is to investigate the level of fatigue these supported gestures produce compared to 2 well known points along the HCI continuum, traditional mid-air gestures and normal keyboard use. The hypothesis is that supported gestures will produce fatigue levels closer to that of keyboard interactions and significantly lower fatigue levels compared to mid-air gestures.

3   Experimental Setup

We conducted a within subject repeated measures experiment across three types of interaction with a video game to examine both the physical and perceived fatigue levels of each. The three conditions were (1) keyboard, (2) supported gestures, and (3) mid-air gestures (Fig. 4).
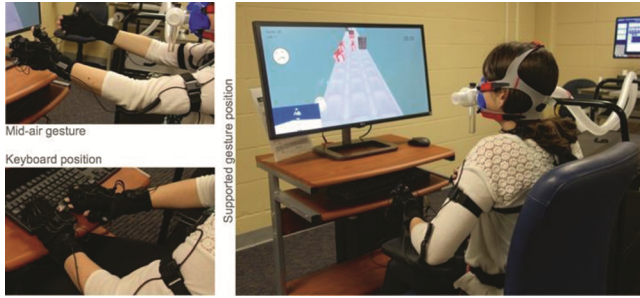
Fig. 4. Experimental setup with each condition demonstrated.

3.1 Participants

The study was conducted with 16 participants from a university population (10 male/6 female). Their age ranged from 20 to 28 with a mean age of 23. The experiment took 2 h to complete per participant.

3.2 Gamification

Games can be used as a prototype to predict human behavior in an interactive system while the system is not fully developed. Our gamified prototypepresented users with a task that would engage and motivate them over the extended time periods they were interacting with the system. Concerning the *degree of functionality* in the field of ergonomics and the *interactivity* [27] associated with authentic product utilization, using a game for the gorilla arm study has the advantages of reducing the users' boredom while increasing their interests in continuously performing the gestures.

Deterding et al. [12, 13] defined the term Gamification which is *"the use of game design elements in non-game contexts"*. The game for the gorilla arm study should satisfy the requirements of (1) motivating users to perform predefined gestures, rather than making them feel they are passively forced to move hands and arms, and (2) being able to easily control how frequent the users should move arms and hands so that it mimics different communication or interaction situations.

We designed a video game called "Happy Ball" and implemented it using the Unity game engine and a high-accuracy motion capturing device (Fig. 5). The goal of the game is to keep the ball happy by avoiding the obstacles and collecting as many presents as possible. The ball can be moved across 3 lanes and is automatically propelled forward at a fixed rate (similar to Temple Run type games). The player controls the left and right movement of the ball to avoid stone obstacles and a stop command to stop bouncing to avoid ice blocks placed above the ball. While the stop command is active, the ball decelerates until it stops moving forward. When the stop command ends, the ball resumes the bouncing motion and accelerates back to the original forward speed. The stop command overrides the controls of left and right movements. For example, when the left hand is performing the stop action, the swiping motion of the right hand will not be able to move the ball to the left lane. After 10 hits to the ball by the obstacles, the

game is over and it automatically restarts after 3 s. During the game, the happiness level (life status) of the ball is represented with a cartoonish facial expression, appearing on the surface of the ball as a texture. There are a total of 10 expressions: elated, joyful, happy, satisfied, neutral, unhappy, depressed, sad, helpless, and crying, each in order corresponding to the number of hits by obstacles in its decreasing order.
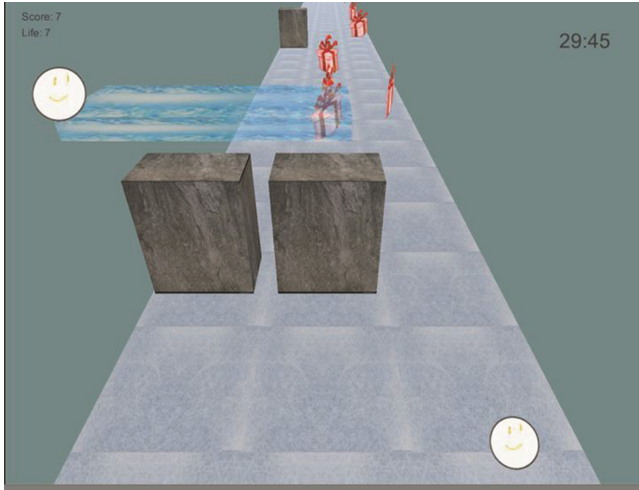
Fig. 5. Screen capture of the "Happy Ball" game. The player controls the ball at the bottom right to collect the presents and avoid the stone and ice obstacles.

3.3 Apparatus

Motion Capture. The participant's gestures were tracked using an IMU based motion capture suit. The configurable motion capture suit from Perception Neuron (https:// neuronmocap.com) was used. Only 2 gloves and the torso strap was used because no movement information was needed for the lower body. The total number of sensors was 23 (9/glove, 1/arm, 3 for torso). Each sensor contains a gyroscope, accelerometer, and magnetometer and wirelessly transmits the motion data to the Axis Neuron Software.

Oxygen Consumption. Oxygen consumption (VO_2) was measured continuously using a TrueOne 2400 metabolic system (ParvoMedics). The TrueOne 2400 is a computerized metabolic system using a gas mixing chamber to analyze the oxygen consumed and carbon dioxide produced. Open-circuit spirometry has been found to provide both reliable [11] and accurate [3] data for the measurement of VO_2. The flow meter and gas analyzers were calibrated prior to each test with a 3L syringe and gases of known concentrations. The participant was fitted with a rubber mask (Hans Rudolph) that covers the nose and mouth, which is connected to the TrueOne 2400 system (Fig. 4).

Perceived Exertion. Subjective physical exertion was measured using the Borg CR10 scale, which is commonly used based on its reliability and validity [6, 20]. This scale

consists of a 12-point scale (0, 0.5, 1–10) with descriptions along 9 of the points ranging from "Nothing at all" to "Impossible". Participants are asked to rate their current level of exertion based on the scale.

3.4   Procedure

A Latin square design was used to counter-balance the order each participant would engage with the 3 conditions. The participant would place the motion capture suit on and go through a 2-minute calibration process. The participant's weight and height were recorded and entered into the TrueOne 2400 software. The participant was then measured for an appropriately sized mask and the mask was placed on the participant, making sure it was an air tight fit. The participant played the Happy Ball game for up to 30 min for each condition while VO_2 was being recorded. The gesture commands are described above for the supported and mid-air gestures. Participants would use the "A" and "D" keys to move left and right and the space bar to stop in the keyboard condition. If the participant experienced excessive fatigue in any condition, they would signal to the experimenter that they would like to stop the current game play. At the conclusion of each condition, they were asked to rate their physical exertion level using the Borg CR10 scale. There was a 5-minute period of inactivity between each condition in order to allow their VO_2 levels to reach their baseline levels again.

4   Results

4.1   Time

The participants had the opportunity to spend up to 30 min engaged with the game for each interaction condition. 27% of participants completed the full 30-minute trial for the mid-air gestures compared to 100% for supported gestures and keyboard. Figure 6 displays the mean time spent across each condition. A paired samples t-test was conducted to compare the time spent in the supported gesture condition and the mid-air
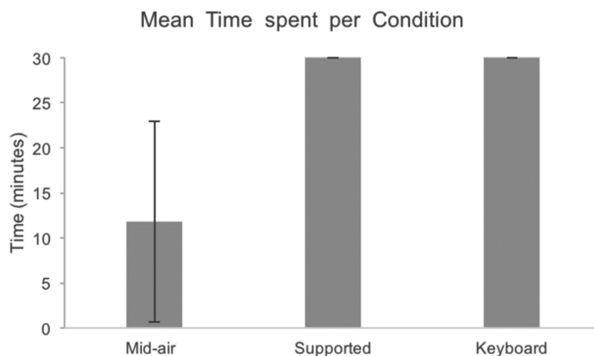
Fig. 6. Mean time spent out of a possible 30 min for each condition. Error bars represent standard deviation.

gesture condition. There was a significant difference in time between the supported gesture conditions ($M = 30.0$, $SD = 0.0$) and the mid-air gestures ($M = 11.85$, $SD = 11.13$); $t(15) = -6.52$, $p < 0.00$ (supported and keyboard results are the same). These results provide strong evidence that confirms the impracticality of mid-air gestures over time given participants could only endure approximately 12 min of this type of activity. On the other hand, participants had no difficulty completing the entire 30-minute time period for both the supported gestures and the keyboard conditions.

4.2 VO$_2$

To test differences of VO$_2$ levels across the 3 conditions, a generalized estimating equations (GEE) with an unstructured correlation matrix was used. GEE is an extension of the general linear model to longitudinal data analysis using quasi-likelihood estimation [1]. VO$_2$ was measured in milliliters per kilogram of body weight per minute (VO$_2$ mL/kg^{-1}/min^{-1}). This model compares the overall VO$_2$ across the conditions. The results indicate a significant difference among the conditions (Wald chi-square $= 77.19$, $p < 0.00$, $N = 47$). Parameter estimates are found in Table 1.

Table 1. Parameter estimates from the GEE model for VO$_2$ mL/kg^{-1}/min^{-1} across the interaction conditions.

Models	β	SE	Wald chi-square	p-value
Intercept	4.77	0.20	592.38	0.00
Keyboard & Mid-air	−1.28	0.15	72.19	0.00
Supported & Mid-air	−1.07	0.12	74.47	0.00

Post-hoc pairwise comparisons of the estimated marginal means were conducted to examine the specific differences. There was a significant difference between supported gestures ($M = 3.69$, $SE = 0.12$) and mid-air gestures ($M = 4.76$, $SE = 0.20$) (Table 2). There was also a significant VO$_2$ difference between the supported gestures and the keyboard ($M = 3.48$, $SE = 0.14$, $p < 0.00$, Table 2) conditions. Lastly, there was a significant difference between the keyboard and the mid-air gesture condition (Table 2). The overall means are displayed in Fig. 7.

Table 2. Multiple comparisons and mean differences in VO$_2$ mL/kg^{-1}/min^{-1} by interaction types.

Comparison	Mean difference	SE	p-value
Supported vs. Mid-air	−1.08	0.12	0.00
Supported vs Keyboard	0.20	0.07	0.00
Mid-air & Keyboard	1.28	0.15	0.00

Mean VO$_2$ Levels per Condition
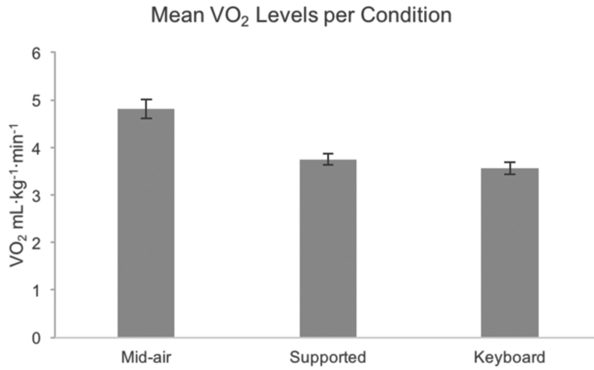
Fig. 7. The mean VO$_2$ in milliliters per kilogram of body weight per minute is presented across each condition. Error bars represent standard error of the mean.

Figure 8 shows the mean VO$_2$ levels across time for each condition. The dotted vertical lines integrate the mean duration performed for each condition reported in the previous section. Even though there was a significant difference between the supported gestures and keyboard activity, these results show how close in physical exertion levels supported gestures are to regular keyboard activity. The level of physical effort for the mid-air gestures, however was comparatively higher to both and produced a 23% increase in oxygen consumption compared to the supported gestures.

Mean VO$_2$ per Condition over Time
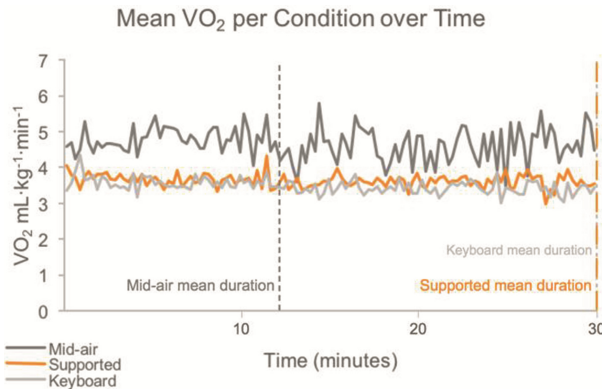
Fig. 8. Mean VO$_2$ for each condition over the 30-minute trial length. The dotted vertical lines represent the mean duration for each condition.

4.3 Perceived Exertion

The perceived physical exertion from each of the conditions was measured by the Borg CR10 on its 12-point scale. A GEE model was also used to analyze the Borg scores across the 3 conditions. The results indicate a significant difference among the conditions (Wald chi-square $= 71.58$, $p < 0.00$, $N = 47$). Parameter estimates are found in Table 3.

Pairwise comparisons of the estimated marginal means were conducted to examine the specific differences. There was a significant difference between supported gestures ($M = 1.70$, $SE = .34$) and mid-air gestures ($M = 5.69$, $SE = 0.57$) (Table 4). There were also significant differences between supported gestures and keyboard use ($M = 1.3$, $SE = 0.34$) as well as the keyboard and mid-air gestures (Table 4). The Borg CR10 label for the mean supported gesture level was "very easy" while the mid-air gesture mean level equates to the exertion level of "Hard" on the scale (Fig. 9). The keyboard mean level falls within the "very, very easy" range (Fig. 9). The perceived exertion results confirm the previous results in that participants estimated that the mid-air gestures required more than 3 times the effort of supported gestures. Even though there was a significant difference between the keyboard and supported hand gestures, both showed very low levels of exertion.

Table 3. Parameter estimates from the GEE model for VO_2 mL/kg/min across the interaction conditions.

Models	β	SE	Wald chi-square	p-value
Intercept	5.69	0.57	98.66	0.00
Keyboard & Mid-air	−4.44	0.52	71.57	0.00
Supported & Mid-air	−3.99	0.52	59.73	0.00

Table 4. Multiple comparisons and mean differences in VO_2 mL/kg/min by interaction types.

Comparison	Mean difference	SE	p-value
Supported vs. Mid-air	−3.99	0.52	0.00
Supported vs Keyboard	0.46	0.21	0.03
Mid-air & Keyboard	4.44	0.52	0.00
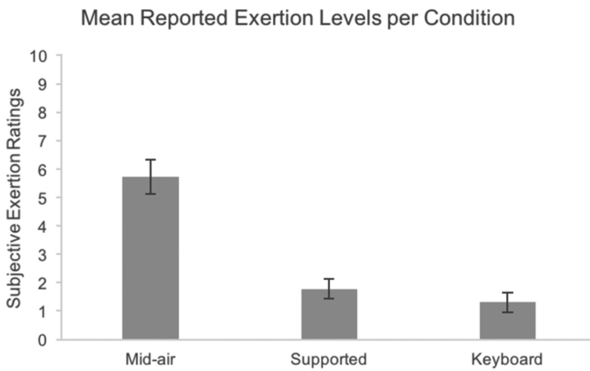
Fig. 9. The mean subjective Borg ratings for physical exertion across conditions (0 = Nothing at all, 10 = Impossible). Error bars represent standard error of the mean.

5   Discussion

The results of this study provided evidence for the hypothesis that the supported gesture interactions produced significantly less fatigue among the participants compared to mid-air type of gestures. Additionally, the fatigue levels from using the supported gestures were very similar to that of traditional keyboard interactions providing positive evidence for sustained use of supported gesture interactions.

The gorilla arm phenomenon has greatly impeded the research and development efforts of gesture based interactions with computer systems. Based on the results of this study, researchers and the technology industry have had good reason to largely ignore the gesture based modality. Only 25% of the participants in this study were able to complete a single 30-minute session using the mid-air gestures, which is only a fraction of the time knowledge workers typically spend in one sitting. The actual physical effort as measured by VO_2 and the perceived exertion level of using mid-air gestures also supports the significant discomfort and fatigue felt by participants.

Any potential modality alternative to the existing status quo of keyboards, mice, and game controllers must be reliable, accurate, and viable for extended periods of use. Mid-air gestures as depicted by the movie industry and made possible by affordable optical motion capture technology such as the Microsoft Kinect and Leap Motion force the user into these mid-air type of gestures. The growing advancements in augmented and virtual reality are renewing interest and some work with gestures as an input modality. However, most of the current efforts ignore the gorilla arm mistakes of the past and appear to be on a path to repeat them. For example, the Leap Motion can be adapted to be used with head mounted displays (HMD) but must be placed on the HMD. With the limited angle of its optical sensors, the user's hands must be raised high enough to be recognized, which brings the user into the high fatigue posture shown in Fig. 1A. Microsoft's augmented reality technology, the HoloLens suffers from the same issue of forcing the user's hands in front of its HMD.

The optical solution for the limited sensing area of a single camera for motion capture has traditionally been a system with several cameras configured around the user. This approach is expensive in terms of both equipment and the dedicated space for this type of setup. These requirements do not allow multi-camera systems to scale well to multiple knowledge workers in a traditional office space. These type of systems also still require an unobscured view of the user's arms and hands, which is fine for mid-air postures (Fig. 1A) but not for supported postures (Fig. 1B) if the user is seated at a desk or table.

The use of IMU based motion capture technology breaks the line-of-sight chains that constrains the optical systems to mid-air type gestures. With IMU motion capture, the user's hands can comfortably rest in supported and natural positions while enjoying a relatively high degree of precision and accuracy for both hand and finger movements.

If supported gestures are to be a viable alternative to keyboard, mouse, and game controller use, users physically must be able perform them for long periods of time without discomfort and fatigue. This study directly compared the level of effort of keyboard use to supported gestures. All the participants from both the keyboard and supported gestures were able to complete the entire 30-minute interaction trial. The results from the participants' physical exertion levels as measured by VO_2 were very

similar. Finally, the participants reported extremely low amounts of perceived exertion for both the keyboard ("very, very easy") and supported gestures ("very easy"), even though they were significantly different. These results suggest that the supported gestures were successful in avoiding the gorilla arm effects that traditional gesture based interactions experience. Overall, the physical effort required to perform supported gestures appears to be very similar to that of using the keyboard.

6   Conclusion and Future Work

The traditional way of implementing gesture based interactions is fundamentally flawed by forcing users to perform arm and hand gestures in mid-air without support or appreciation of natural resting positions. In this paper, we have traced one potential cause of mid-air gestures to the constraints that come with the use of optical motion capture systems. We avoided this constraint by implementing a novel vector based gesture recognition approach using an IMU motion capture system. The newly designed supported gestures were found to require similar levels of physical and perceived exertion as keyboard use and avoided the gorilla arm effects experienced by participants during their mid-air gesture interactions. Some of the limitations of this study include a very limited gesture vocabulary and data collection over only a 30-minute duration. Given the promising results shown with supported gestures, our future efforts will be to continue developing the supported gesture vocabulary and examine their integration with other input modalities.

References

1. Ballinger, G.A.: Using generalized estimating equations for longitudinal data analysis. Organ. Res. Methods **7**(2), 127–150 (2004)
2. Badi, H.S., Hussein, S.: Hand posture and gesture recognition technology. Neural Comput. Appl. **25**(3), 871–878 (2014)
3. Bassett Jr., D.R., Howley, E.T., Thompson, D.L., King, G.A., Strath, S.J., McLaughlin, J.E., Parr, B.B.: Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. J. Appl. Physiol. **91**, 218–224 (2001)
4. Bergqvist, U., Wolgast, E., Nilsson, B., Voss, M.: Musculoskeletal disorders among VDT workers: Individual, ergonomic, and work organizational factors. Ergonomics **38**(4), 763–776 (1995)
5. Bolt, R.A.: "Put-that-there": voice and gesture at the graphics interface. In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1980), pp. 262–270 (1980)
6. Borg, E., Kaijser, L.: A comparison between three rating scales for perceived exertion and two different work tests. Scand. J. Med. Sci. Sports **16**(1), 57–69 (2005)
7. Boring, S., Jurmu, M., Butz, A.: Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays. In: Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design (OZCHI 2009), pp. 161–168 (2009)
8. Brown, M.A., Stuerzlinger, W., Mendonca Filho, E.J.: The performance of un-instrumented in-air pointing. In: Proceedings of Graphics Interface Conference, pp. 59–66 (2014)

9. Carmody, T.: Why 'gorilla arm syndroome' rules out multitouch notebook displays. Wired **10,** October 2010

10. Cadoz, C.: Les Realites Virtuelles. Flammarion, Dominos (1994)

11. Crouter, S.E., Antczak, A., Hudak, J.R., Della Valle, D.M., Haas, J.D.: Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. Eur. J. Appl. Physiol. **98**, 139–151 (2006)

12. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9–15 (2011)

13. Deterding, S., Sicart, M., Nacke, L., O'Hara, K., Dixon, D.: Gamification using game-design elements in non-gaming contexts. In CHI 2011 Extended Abstracts on Human Factors in Computing Systems, pp. 2425–2428 (2011)

14. Van Edwards, V.: 5 secrets of a successful TED talk. http://www.scienceofpeople.com/2015/03/secrets-of-a-successful-ted-talk/. Accessed 14 Sep 2016

15. Freeman, D., Vennelakanti, R., Madhvanath, S.: Freehand pose-based gestural interaction: Studies and implications for interface design. In: IEEE Proceedings of 4th International Conference on Intelligent Human Computer Interaction, pp. 27–29 (2012)

16. Freivalds, A.: Biomechanics of the Upper Limbs: Mechanics, Modeling and Musculoskeletal Injuries, 2nd edn. CRC Press, Boca Raton

17. Guinness, D., Jude, A., Michael Poor, G., Dover, A.: Models for rested touchless gestural interaction. In: Proceedings of the 3rd ACM Symposium on Spatial User Interaction, pp. 34–43 (2015)

18. Guna, J., Jakus, G., Pogacnik, M., Tomazic, S., Sodnik, J.: An analysis of the precision and reliability of the Leap Motion sensor and its suitability for static and dynamic tracking. Sensors **14**(2), 3702–3720 (2014)

19. Hasan, H., Abdul-Kareem, S.: Static hand gesture recognition using neural networks. Artif. Intell. Rev. **41**(2), 147–181 (2014)

20. Hincapié-Ramos, J.D., Guo, X., Moghadasian, P., Irani, P.: Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1063–1072 (2014)

21. Kendon, A.: Gesture: Visible Action as Utterance. Cambridge University Press, New York (2004)

22. Andrew Life, M., Pheasant, S.: An integrated approach to the study of posture in keyboard operation. Appl. Ergon. **15**(2), 83–90 (1984)

23. Luzhnica, G., Simon, J., Lex, E., Pammer, V.: A sliding window approach to natural hand gesture recognition using a custom data glove. In: 2016 IEEE Symposium on 3D User Interfaces (3DUI), pp. 81–90 (2016)

24. Nielsen, J.: The usability engineering life cycle. IEEE Comput. **25**(3), 12–22 (1992)

25. Nielsen, M., Störring, M., Moeslund, T.B., Granum, E.: A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 409–420. Springer, Heidelberg (2004). doi:10.1007/978-3-540-24598-8_38

26. Novack, M., Goldin-Meadow, S.: Learning from gesture: how our hands change our minds. Educ. Pscyhol. Rev. **27**, 405–412 (2015)

27. Sauer, J., Sonderegger, A.: The influence of prototype fidelity and aesthetics of design in usability tests: effects on user behaviour, subjective evaluation and emotion. Appl. Ergon. **40**(4), 670–677 (2009)

28. Segen, J., Kumar, S.: Look ma, no mouse! Commun. ACM **43**(7), 102–109 (2000)

29. Trigueiros, P., Ribeiro, F., Reis, L.P.: A comparison of machine learning algorithms applied to hand gesture recognition. In: 7th Iberian Conference on Information Systems and Technologies, Madrid, pp. 1–6 (2012)
30. Underkoffler, J.S.: The I/O Bulb and the Luminous Room. Ph.D. Dissertation. Massachusetts Institute of Technology, Cambridge (1999)