Virtual Exertions: a user interface combining visual information, kinesthetics and biofeedback for virtual object manipulation

Kevin Ponto* Ryan Kimmel† Joe Kohlmann‡ Aaron Bartholomew§ Robert G. Radwin¶

University of Wisconsin–Madison

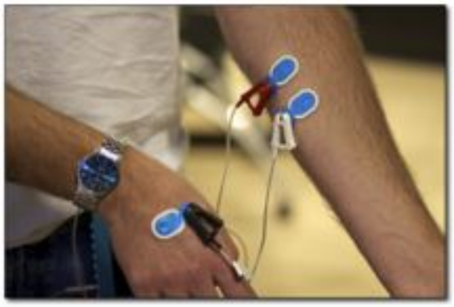

Figure 1: Using visual information, kinesthetics and biofeedback from electromyograms (EMG), users can grasp, move and drop virtual objects.

Abstract

Virtual Reality environments have the ability to present users with rich visual representations of simulated environments. However, means to interact with these types of illusions are generally unnatural in the sense that they do not match the methods humans use to grasp and move objects in the physical world. We demonstrate a system that enables users to interact with virtual objects with natural body movements by combining visual information, kinesthetics and biofeedback from electromyograms (EMG). Our method allows virtual objects to be grasped, moved and dropped through muscle exertion classification based on physical world masses. We show that users can consistently reproduce these calibrated exertions, allowing them to interface with objects in a novel way.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented and virtual realities H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies

1 Introduction

Virtual reality environments utilize immersive experiences to induce a feeling of presence [15]. While advancements in resolution and refresh rate may add to the immersive capabilities of a virtual reality system, they may not strengthen the user's sense of presence. For instance, a study by Slater, et al. found that presence is enhanced when interaction techniques are employed that permit the user to engage in whole-body movement [16]. Barfield and Hendrix reported that the level of interactivity between the subject and the virtual environment, rather than fidelity of the visual scene, was related to the perception of presence [2]. In this sense, a virtual environment in which a person can interact naturally, intuitively and instinctively will enhance the user's illusion of presence [4].

*e-mail: kponto@cs.wisc.edu
†e-mail: rkimmel@wisc.edu
‡e-mail: jkohlmann@wisc.edu
§e-mail: abartholome2@wisc.edu
¶e-mail: radwin@bme.wisc.edu

It is therefore beneficial for users to grasp, hold and manipulate objects in a virtual world as they do in the physical world. Determining if the user is attempting to grasp an object is a more difficult task than simple collision detection [17]. Common grasping triggers include button presses, hand gesture commands and speech commands [11]. Hand pose gesture recognition attempts to match natural interaction, but humans use their hands in many different ways to pick up objects, not all of which are recognizable with hand gesture systems, as shown by Zachmann [17].

We present the idea of virtual exertions, a method utilizing biofeedback from electromyograms (EMG) along with visual and kinesthetic information to manipulate virtual objects. Virtual exertions are physical interactions with immersive virtual objects acted on through body motions and muscle contractions, mimicking exertions against physical objects. Users control virtual objects with hand and body movements and muscle contractions similar to those used on physical objects.

2 Previous Work

Bowman and Hodges evaluated grasping and manipulation techniques, including virtual representations of arms and hands and ray casting, in virtual environments [3]. Schlattmann et al. provided a summary of interaction techniques for markerless handtracking [14]. For much of this work, users were required to fit their hand to a grasping pose to acquire an object, as no information of the exertion forces could be ascertained [17]. Exertion force studies generally require fixed position input devices [7, 10]. These devices can provide haptic feedback, but their lack of mobility reduces the user's level of interactivity and immersion.

Researchers have used EMG sensors for human computer interaction, such as Costanza et al., who created intimate user experiences by analyzing subtle movements [6]. Saponas et al. [13] explored methods to classify finger gestures using a muscle sensing armband. Saponas et al. also used forearm EMG to classify finger gestures on physical surfaces, allowing them to interpret four-finger gestures with high degree of accuracy [12]. These studies focused on gesture recognition and classification of physical actions using EMG signals; we are interested in analyzing exertions as an interface for virtual environments.

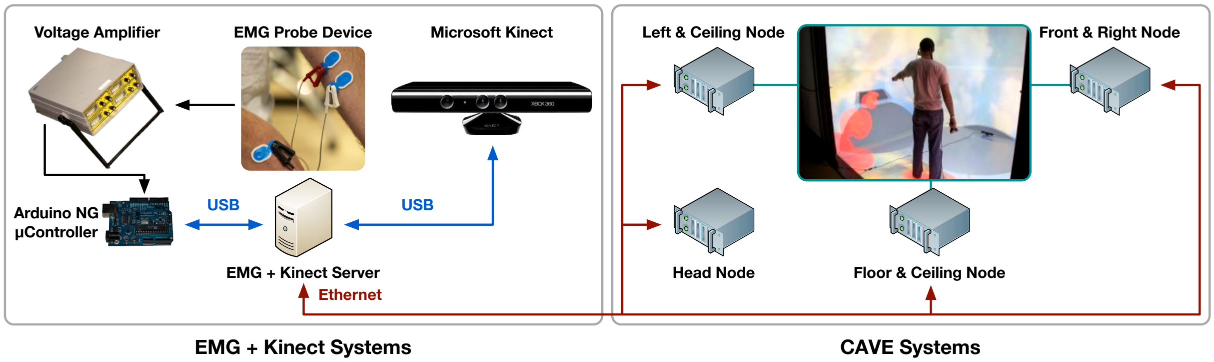

Figure 2: System architecture. The information collected by the EMG and Kinect Server is sent to the CAVE system via TCP. The inset photo (right) shows the aligned Kinect skeletal joints, shown in red for debugging, which indicate the user's joint positions within the virtual environment.

3 Method

Our goal is to create an interface in which virtual objects react to hand and body movements and contraction of muscles, similar to the physical world. By using kinesthetic information and biofeedback, users have the ability to grasp, lift, move and drop objects. Unlike previous work, our method gives virtual objects the illusion of mass by requiring users to exert a calibrated amount of exertion to grasp and hold them. Our system uses an EMG system, a Microsoft Kinect and a Virtual Reality environment (Figure 2).

3.1 Biofeedback System

Virtual interactions, such as picking up objects, are defined by the level of muscle activity required to manipulate corresponding physical objects. Our approach uses surface EMG to record the muscle activity of the flexor capri ulnaris muscle. Although no forces are exerted by the hands and body appendages, muscle activity in the users forearm mimics the intensity of exertions made when acting against physical objects. Brown and McGill [5] observed a linear relationship in the EMG moment relationship of trunk muscles when measuring antagonist muscle co-activation.

The belly of the flexor capri ulnaris muscle was first located while a subject performed an isometric contraction holding a 4.5 kg load. Electrode positioning was performed according to the guidelines proposed by Mogk and Keir [9]. Electrode positioning over the muscles was confirmed by palpation and signal response during specific exertions [1, 9]. The skin was cleaned with 91% isopropyl alcohol and allowed to dry for 1 minute. Silver-silver chloride electrodes were located over the muscle belly, parallel to the muscle fibers, with an inter-electrode distance of 2.5 cm. A reference electrode was placed on the dorsal side of the opposite hand, away from the electrically active area.

The surface EMG signals were amplified, integrated (IEMG), converted and sampled using an analog-digital converter connected to an Arduino NG microcontroller [8]. The IEMG signals are directly proportional to overall muscle activity and consequently to forces biomechanically linked to the limbs and torso [5]. The IEMG signals were calibrated using a series of exertions in the postures assumed when performing the task to be mimicked in the virtual environment.

As IEMG signals are very small in magnitude, they generally contain large amounts of noise that can make exertion classification difficult. Kalman filters estimate "true" values by predicting a value, estimating the uncertainty and computing a weighted average of the predicted value and the measured value. We apply a constant velocity Kalman filter in which we model the IEMG signal value and the derivative for the state variables. We determined all parameter values for our filter through empirical observations (for all data shown in this paper, filter parameters $\sigma_a^2 = .004$, R = $1E^{-5}$and $\Delta t = 0.033$).

We selected peak-average value rather than a simple mean as it tended to produce a more representative baseline. This produces a better segmentation of action and non-action.

To calibrate each user we monitored the IEMG signals while holding calibration objects, consisting of a masses weighing 0.74 kg (1.63 lbs), 1.13 kg (2.5 lbs), 1.36 kg (3 lbs), 2.27 kg (5 lbs) and 4.54 kg (10 lbs). The user was asked to grasp each object, hold it for five seconds and then release.

As shown by Brown and McGill [5], the amount of exertion force scales linearly with mass of the object. This simple linear conversion from mass to exertion force enables each virtual object to be assigned a Minimum Exertion Force (MEF) value. This MEF value represents the amount of force required to pick up and hold an object. Generally the linear fit equations had R^2 values greater than 0.99 in our testing. The method used to grasp, move and drop objects is described in Section 3.3.

3.2 Kinesthetic System

We used a Microsoft Kinect system as a low cost and unobtrusive means to capture the user's posture information and skeleton using Microsoft's Kinect SDK Beta 2 (released November 1, 2011). As the Kinect SDK operates in its own reference frame, the positions of each skeletal joint is given as a distance from the Kinect camera. To use this information, we need to convert from the Kinect reference frame into the virtual reference frame.

We first align the Kinect system with the front wall of the CAVE to remove rotational discrepancies for the yaw and roll axis. There are still rotational discrepancies in the pitch axis, however (Figure 3). To correct for these, we calculate the Kinect's "up" direction by asking the user to stand straight up and recording the position of their joints. From this we create a vector from the center of the hip to the center of the shoulder that represents the user's "up" direction. We can then calculate the pitch rotational discrepancy (θ).

To create our correction matrix, we first translate the Kinect joints relative to the head joint position (I, J, K), rotate about the x-axis (θ) and finally translate the joints in the virtual world to match the virtual worlds head location (X, Y, Z). We require one more correction—while the Kinect system locates the center of the head, the head tracking system locates the users eyes, several centimeters above. We compensate for this by applying a small offset in the z direction (δ).

3.3 Virtual Reality Environment

The presented method is designed to work in a Cave Automatic Virtual Environment (CAVE). For the methods involved in this paper, it is necessary for the users to have the illusion that their hand can grasp virtual objects. To achieve this effect, users must be head

Figure 3: Comparing the uncorrected skeleton (left) with the skeleton after using the correction matrix (right). The skeleton joints, shown in red for debugging, are positioned in front of the viewer for the uncorrected version, while the joints are aligned with the viewer's perspective in the corrected version.

tracked and be receiving stereo 3D visual information. Other immersive display technologies, such as head mounted displays, provide another way of generating these kinds of immersive experiences.

Our method for grasping objects is similar to that of Zachmann [17]. First, we determine if there has been a collision between the virtual object and the hand, as detected by the physics engine. We represent the hand with an invisible sphere, 5 cm in radius. If there is a collision between sphere and object, we compare the user's exertion force against the object's MEF value. If the force being applied is greater than the MEF value, the object is considered to be grasped, and its movement follows the user's hand. If the user's exertion force falls below the MEF value, the object is dropped.

4 Results

We created several different environments to test our system. The first environment consisted of two circular tables in which users were tasked with moving objects from one table to the other. The objects consisted of books and dumbbells, mirroring the physical objects the user can train with. We also utilized recreations of kitchen and bathroom environments filled with everyday objects, such as toothbrushes, deodorant, soap, teapots, pans and cups that the user could manipulate.

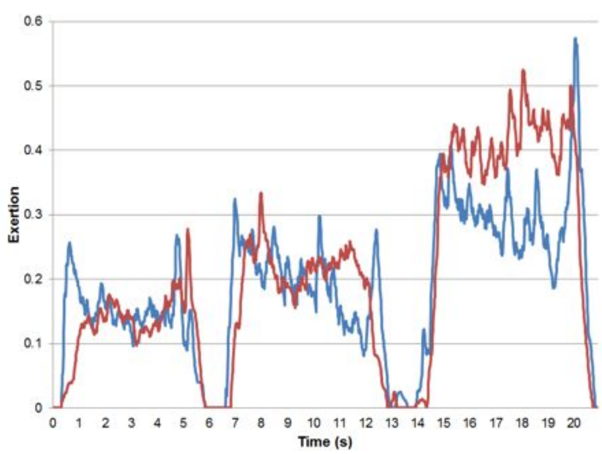

Figure 4: A demonstration of the user's exertion force while lifting a physical object, shown in blue, and a virtual object, shown in red. Left to right, the graph shows the exertion force from lifting 0.74 kg (1.63 lbs), 1.36 kg (3 lbs) and 2.27 kg (5 lbs) objects, respectively.

For the system to be effective, users needed to be able to virtually match the exertion that they would normally need to produce for lifting a physical object of equal mass. To test this, we had users lift objects of 0.74 kg (1.625 lbs), 1.36 kg (3 lbs) and 2.27 kg (5 lbs), first physically and then virtually. Figure 4 shows a graph of these exertions for a user, with physical exertions shown in red and virtual exertions shown in blue. In general, users were able to generate a similar force for the virtual object, just as they would have used to lift the physical object.

It was also important to test the responsiveness of our system to see if it properly reflected the user's intentions. To accomplish this, we compared our novel interaction method to a more traditional one. When the user intended to grasp a virtual object, they were asked to push a trigger button on a wireless controller. When the user wanted to drop the object, they were instructed to release the button. A total of four participants completed 10 trials for each weight. Table 1 shows the average time differences from all trials.

Six individuals familiar with the CAVE assisted in the development of this system. Most users were able to pick up objects without instruction after going through the calibration procedure. Users' statements about the system were generally positive, stating that they found the method of interaction easy and engaging. Some users enjoyed the ability to manipulate objects either with or without a physical hand gesture. The major complaint was the prep work and movement limitations imposed by the wired EMG device.

Method	0.74 kg	1.36 kg	2.27 kg
Grasp	20 ms (SD=44)	309 ms (SD=70)	359 ms (SD=100)
Release	32 ms (SD=100)	90 ms (SD=40)	5 ms (SD=5)

Table 1: The Grasp row shows the difference in time between when the user pressed the controller button and when our system classified the object's state as grasped. The Release row shows the difference in time between when the user pressed the controller button and when our system classified the object's state as being dropped.

Our method was most effective for lightweight virtual objects ($<$ 1.13 kg (3 lbs)). Users were able to grasp these objects with little latency compared to a simple button press. Dropping these objects generally worked effectively, but sometimes users flexed their hands in the process, thus increasing their exertion for a short interval. This in turn created the appearance that objects were stuck to the hand for short periods of time. For objects of greater mass, the grasping stage was much less precise, resulting in a higher standard deviation. This was due to the time needed for the user to reach the correct level of exertion. Dropping virtual objects with a greater MEF value was generally easier for users, with very heavy objects being very low latency compared to a simple button release.

5 Discussion

While this system provides an initial proof of concept, further testing is still required. As this method uses the human body, a study of effectiveness over a broad population would be insightful. For instance, an understanding of how fatigue is incurred through only muscle tension, i.e. without an opposing force, would be necessary to make this interface more general purpose. Furthermore, a better understanding of the user's sense of proprioception while working in a virtual environment would enable new approaches that increase the precision of object acquisition and placement.

Natural interactions are not necessarily the most efficient way to accomplish a task. For instance, users experienced fatigue after extended periods of time while using the system, as they would naturally. Existing input systems such as game controllers and wands are clearly more efficient in this regard, but they do not reflect common, naturally occurring sensorimotor contingencies. That is, a button press is an efficient, though unnatural interaction without an apparent causal relationship between the user and the object.

The current system only distinguishes interactions in a binary way—if an item isn't "grasped", it is immediately "dropped". Because the user is quick to react to the sight of an object beginning to fall, they compensate by increasing their exertion level. Thanks to this behavior, objects were rarely dropped in practice. Haptic interfaces would give the user additional feedback beyond vision alone, providing perceptual cues about a virtual object's MEF.

As shown in [3], objects can also be acquired via ray-casting, allowing users to manipulate objects from a distance. We implemented this interaction paradigm by creating a ray from the user's elbow joint pointed to the user's hand. This allowed users to point their arm towards an object and exert their muscles to grab the object from a distance. This small enhancement offers a means to increase the efficiency of this natural interaction.

Our prototype system is also somewhat cumbersome in its current state. While the wires attached to the user's body are long enough to enable full traversal of the CAVE, users can still feel restricted. Wireless armband EMG devices have been prototyped by both Costanza et al. [6] and Saponas et al. [13]. These types of devices not only remove the wired connection, but also require much less preparatory work.

The user's movement is also somewhat restricted by the single Kinect camera's field of view. We believe that by adding and registering multiple Kinect camera devices, these restrictions could be greatly reduced. The system is also limited by the skeletal construction provided by the current version of the Microsoft Kinect SDK. While this SDK cannot capture hand poses, future research may be able to accomplish this with more sophisticated hardware or computer vision techniques.

Future work may benefit from a more complex classification system that could enable classification of not only muscle exertion, but also of the user's gestures and actions. Enhancements may provide a means of disambiguating gestures that appear similar based purely on movement. Also, as the force generated by the user is not digital in nature, adding classification may give extra insight into the user's actions. For instance, it may be possible to differentiate dropping from flicking and throwing. Classification could also be used to differentiate the way in which the user has exerted their muscles.

As we only use the forearm muscles for our biofeedback system, it is most useful for capturing gripping actions. For lifting cradled objects, the primary muscle group used for lifting may switch to the bicep muscles. Thusly, in future work, it may be important to focus on multiple muscle groups. This would, in its simplest form, enable users to lift objects with multiple hands. One might also use virtual exertions as a form of physical rehabilitation for a patient recovering from surgery. Furthermore, an application of this system could utilize the user's leg muscle exertions and overall body position to naturally traverse a virtual environment simply by "walking".

6 Conclusion

This paper presents a novel interface for virtual environments by combining kinesthetic information and biofeedback from electromyograms (EMG). This method more closely matches the way people naturally interact with physical objects through grasping, moving and dropping without need for buttons, hand gestures or speech commands. This gives virtual objects the illusion of mass by requiring users to exert a calibrated amount of force to grasp and hold virtual objects. Users were consistently able to reproduce these calibrated exertions when manipulating virtual objects of varying mass. Future work will focus on making the system more natural than the current setup and on further enhancing its overall effectiveness.

7 Acknowledgements

We would like to acknowledge the support of the the Living Environments Laboratory, the Wisconsin Institute of Discovery, and the CIBM training program. This project was supported in part by NLM award 5T15LM007359. We would specifically like to thank Patricia Brennan, Hyun Joon Shin, Kendra Jacobsen, Andrew Cherry and Michelle Craft for their support and assistance.

References

[1] O. Alizadehkhaiyat, A. C. Fisher, G. J. Kemp, K. Vishwanathan, and S. P. Frostick. Upper limb muscle imbalance in tennis elbow: A functional and electromyographic assessment. *Journal of Orthopaedic Research*, 25(12):1651–1657, 2007.

[2] W. Barfield and C. Hendrix. The effect of update rate on the sense of presence within virtual environments. *Virtual Reality*, 1:3–15, 1995. 10.1007/BF02009709.

[3] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, I3D '97, pages 35–ff., New York, NY, USA, 1997. ACM.

[4] J. Bragt. Do exertion interfaces provide better exercise environments. In *2nd Twente Student Conference on IT, The Netherlands*, volume 24. Citeseer, 2005.

[5] S. H. Brown and S. M. McGill. Co-activation alters the linear versus non-linear impression of the emgtorque relationship of trunk muscles. *Journal of Biomechanics*, 41(3):491 – 497, 2008.

[6] E. Costanza, S. A. Inverso, R. Allen, and P. Maes. Intimate interfaces in action: assessing the usability and subtlety of emg-based motionless gestures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 819–828, New York, NY, USA, 2007. ACM.

[7] M. Ferre, I. Galiana, J. Barrio, P. García-Robledo, A. Giménez, and J. López. Two-hand virtual object manipulation based on networked architecture. In *Proceedings of the 2010 international conference on Haptics - generating and perceiving tangible sensations: Part II*, EuroHaptics'10, pages 130–135, Berlin, Heidelberg, 2010. Springer-Verlag.

[8] A. Gibb. *New media art, design, and the Arduino microcontroller: A malleable tool*. PhD thesis, Pratt Institute, 2010.

[9] J. P. Mogk and P. J. Keir. Crosstalk in surface electromyography of the proximal forearm during gripping tasks. *Journal of Electromyography and Kinesiology*, 13(1):63 – 71, 2003.

[10] M. Monroy, M. Oyarzabal, M. Ferre, A. Campos, and J. Barrio. Masterfinger: Multi-finger haptic interface for collaborative environments. In *Proceedings of the 6th international conference on Haptics: Perception, Devices and Scenarios*, EuroHaptics '08, pages 411–419, Berlin, Heidelberg, 2008. Springer-Verlag.

[11] S. Oviatt, A. DeAngeli, and K. Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. In *Referring Phenomena in a Multimedia Context and their Computational Treatment*, ReferringPhenomena '97, pages 1–13, Stroudsburg, PA, USA, 1997. Association for Computational Linguistics.

[12] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 167–176, New York, NY, USA, 2009. ACM.

[13] T. S. Saponas, D. S. Tan, D. Morris, J. Turner, and J. A. Landay. Making muscle-computer interfaces more practical. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 851–854, New York, NY, USA, 2010. ACM.

[14] M. Schlattmann, T. Nakorn, and R. Klein. 3d interaction techniques for 6 dof markerless hand-tracking. In *International Conference on Computer Graphics, Visualization and Computer Vision (WSCG'09)*, 2009.

[15] M. Slater. Measuring presence: A response to the witmer and singer presence questionnaire. *Presence*, 8(5):560–565, 1999.

[16] M. Slater, A. Steed, J. McCarthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):469–477, 1998.

[17] D. Zachmann. *Virtual Reality in Assembly SimulationCollision Detection, Simulation Algorithms, and Interaction Techniques*. PhD thesis, TU Darmstadt, 2000.