Real World and Virtual Reality Interactions

Alif Jakir* Robert Miller†

Clarkson University, Potsdam, NY

ABSTRACT

This study explores the behavioral dynamics of user interactions within virtual reality (VR) environments compared to real-world settings, aiming to enhance the efficacy of VR as a training tool. Employing a combination of quantitative measures and user feedback, we conducted a detailed analysis of user movements and task execution across diverse tasks involving object manipulation in both VR and real-world scenarios. The research utilized advanced motion tracking technologies to capture and compare subtle differences in user behavior and interaction patterns. Findings from this study reveal significant variations in user performance, highlighting key areas where VR simulations diverge from real-world interactions. These discrepancies are particularly pronounced in the precision and timing of movements, with implications for the design of more accurate and responsive VR training programs. The insights gained contribute to a deeper understanding of the potential and limitations of VR technology in replicating real-world experiences, offering valuable guidance for developers and researchers in enhancing the realism and effectiveness of VR training environments.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 INTRODUCTION

Virtual reality (VR) is expanding into a variety of fields as a teaching tool. One type of educational application focuses on providing users an immersive environment and the ability to perform life-like interactions to proceed through a task. Environments such as these provide a low-cost and low-risk way to achieve hands on experience for tasks where those values are otherwise not possible. For these environments to be an effective learning tool it is important to understand the differences between interactions in the simulated environments and the real-world.

Advancements in computing hardware and rendering techniques allow VR interactions to be seen at highly immersive and realistic levels. Additional advancements in the study of haptic feedback devices increase the level of presence and distinction in interactions. While both of these slice the veil between how a subject perceives real and virtual interactions, variations between the underlying behaviors remain under-explored.

If VR environments are to be used as educational simulations to provide low-risk training for high-risk real-world situations, understanding how interactions vary between the settings is paramount. Aiding in the study of this we provide:

- The first dataset that consists of VR interactions with corresponding real-world interactions.
- Analysis of the dataset for within and cross modality consistency of action performance.

*e-mail: jakirab@clarkson.edu

†e-mail: romille@clarkson.edu

In this work we show that users show measurable variations when moving different objects in similar patterns in a VR environment. We show that users have a larger/smaller/similar variance when performing the same task in a real-world setting using the same objects and patterns. In the cross-modal setting, we observe that subjects perform the tasks in a similar/divergent manner in the two settings.

2 RELATED WORK

2.1 VR Behavioral Biometrics

Behaviors of subjects interacting in VR as a means to perform authentication has been studied extensively [6–10, 12]. Initial approaches used movement patterns recorded from a head-mounted device [6, 9, 12]. Since the VR systems used had no hand controllers, the interactions recorded was generally passive, reactions to music [6], presentations [12], and targets [9]. Later systems incorporated hand controllers which enabled a wider range of interactions. Interactions enabled by hand controllers include picking up objects, placing objects, pointing, throwing objects, as well as object specific interactions like pulling the string of a bow back. The inclusion of hand controllers with VR systems allows for more complex tasks to be used for behavioral authentication [7, 8, 10].

Olade et al. [10] looked at subjects picking up, rotating, and placing objects. Using the position and orientation of the controllers and headset along with eye location, they achieve a highest accuracy of 98.6% using k nearest neighbors. Liebers et al. [7] studied authentication using an archery task and a bowling task. Using recurrent neural networks they achieved accuracies up to 90% for the archery task. Miller et al. [8] focused on cross system authentication. Here subjects picked up and threw a ball at a target using three different VR systems. Cross-system accuracies range from 87.82% to 98.53% depending on which system pairing was used.

While VR behavioral authentication looks at how subjects perform interactions, it does so primarily through the lens of authentication. This shifts the focus on behavior similarity from being absolute, how consistently does one user perform an interaction, to a relative view, a comparison between how similar one user's performance is with their other performances and with the performances of other users. Additionally, the studies largely look at data recorded from using a single VR system. Only Miller et al. [8] look at cross-system authentication. Cross-system comparisons have similarities with our work. The tracking system and form factor vary for each system. In VR the interaction of picking up and throwing a ball relies on a controller acting as an interface. This might mean certain poses in VR feel unnatural but are natural in the real-world. Additionally, the physical and tactile details of performing the task might cause divergences in behavior based on where and how it is comfortable to pick up the object.

2.2 VR Haptic Devices

Haptic feedback can act as an additional layer of immersion to VR as it allows for the sensation of touch in a virtual space, which helps to create a more life-like virtual experience. Touch feedback increases the action-confirmation loop, reduces cognitive load, and enhances the illusion of presence. A sense of touch is therefore fundamental to a feeling of being physical instantiated within a virtual space. At-market non-enthusiast devices offer relatively

rudimentary haptics often only simple vibration. The combination of haptics into a multisensory feedback paradigm has the capability to provide richer virtual environment information. Several recent haptic devices explore their use in virtual tasks [9-12].

Pezent et al. [11] designed a device, **Tasbi** (**T**actile **A**nd **S**queeze **B**racelet **I**nterface) which is capable of generating squeeze and vibrotactile haptic feedback. Their method is capable of approximating or substituting sensations in hands, as their device creates evenly distributed normal squeeze forces around the wrist, with consistent position and skin coupling. Haptics for wrists offer advantages such as freeing up hands, which allow users to manipulate the physical world without hindrance. The actual configuration of tactile generation is also a factor for increasing task accuracy guided with haptics. They were able to conclusively discover that varying visual and haptic modalities together produced more accurate responses than those modalities alone.

Zenner et al. [15] modified controllers in order to create haptic feedback based on drag and weight shifts. The controller is able to dynamically adjust its surface area in order to leverage air resistance and utilize a changing mass distribution. This allows for the conveyance of resistance and inertia in a virtual task, increasing haptic realism. Based on their user study, they found their *Drag:on* system was able to provide distinguishable levels of haptic feedback to subjects. Building on these types of interfaces can allow for rendering various virtual mechanical resistances, virtual gas streams, and virtual objects with differing scales and materials.

Zenner et al. [16] looked at a novel combination of **Dynamic passive Haptic Feedback (DPHF)** with **Haptic Retargeting (HR)**. **DPHF** is a hardware-based technique that allows for dynamic adjustment of physical properties of an object at runtime, in order to emulate the dynamical properties of various virtual objects. **HR** is a software-based technique that manipulates virtual hand rendering in order to control real hand movement. The authors combine these techniques for weight shifts in a virtual stick. They are able to validate that this combination of techniques creates a greater *Similarity* metric, showing users overestimate weight shifts in virtual spaces.

Haptics try to aid subjects performing tasks by providing additional feedback. The metrics used reflect this, generally focusing on subject perception when using a particular device. Pezent et al. [11] had subjects push buttons of varying firmnesses and looked at how well subjects could discern the firmness level based on their wrist mechanism. It is likely that differing haptic and input devices will yield varying levels of verisimilitudes between real and virtual behaviors. Our work looks to illuminate behavior differences between real-world and VR environments by studying underlying behavior changes in addition to the perception a subject has on their performance.

2.3 Gait Recognition

A related field studying behavior is gait recognition. Gait, how people walk, is used as an identifying factor. The data used in these studies varies being vision-based, images and videos, or wearable-based, inertial data. These two modalities correspond with the modalities in our real-world VR comparison with the real interactions being visual-based and the VR interactions "wearable-based". Vision-based data sees usage in two ways, silhouettes and skeletons. Using silhouettes alone is most common, followed by skeletons alone, then using a combination of the two [13]. Additionally, gait recognition primarily does not use multiple modalities in an adversarial manner. Instead it seeks to use both forms of data together to further understanding of a subjects movements.

Bhalla et al. [1] provided an approach for activity recognition using gait. They use data from a doppler and an IMU. The data of one sensor, the IMU, acts as a training aid only. The addition of the IMU samples helps in learning the latent feature representation for the doppler with limited data. Delgado-Escaño et al. [3] looked at overcoming missing modalities for multi-modal gait recognition by structuring a deep learning architecture to encode each modality separately and then apply a gate and merge operation combining the encodings. Hoang et al. [5] look at gait recognition using data recorded using two different sensors in a comparative way. Here the two sensors are bundled together before recording to study solely authentication rates from data recorded across the sensors.

Figure 1: The six objects used throughout the study. Starting on the left moving right: box of macaroni and cheese, ice cube tray, applesauce cup, vase, water bottle, isopropyl alcohol container.

While gait looks at behaviors, multi-modal approaches tend towards sensor fusion using all modalities at once in a cooperative manner. Our work looks at the data recorded in different modalities comparatively. The real to VR comparison faces challenges due to an underlying behavior shift that is not present in gait studies. The real interactions have a level of tactile detail unable to be matched in VR which will affect behaviors. Additionally there are differences present in the data used. While skeleton data is used in gait, our VR and real "skeletons" feature fewer joints using only those based on the controllers and headset.

2.4 Cross-Modal Person Re-identification

A sub-field of person re-identification looks to match RGB images with IR images. This arises out of the need to identify people in poorly lit scenarios. RGB to RGB identification faces challenges due to large variations, clothing, camera orientation, occlusion, and lighting, for a particular subject. RGB to IR identification faces the same challenges with an increase in difficulty due to data modality differences.

Wang et al. [14] used a generative adversarial network (GAN) to generate a fake IR image from an RGB image. The fake IR image is used with the real IR image to authenticate the subject. The GAN model they utilize attempts to bridge the cross-modal gap both in pixel space, by generating fake IR images, and in feature space, by using the features extracted from both fake and real IR images as part of the discriminator. Hao et al. [4] attempt to align learned features by introducing additional loss functions. One loss function seeks to correlate the RGB images and the IR images in the feature space. The other loss function stems from the assumption that the modalities are two forms of the same information and that high level representations of each should have similar distributions. Choi et al. [2] use a deep learning based approach to disentangle ID-discriminative features, body shape and clothes pattern, from ID-excluded features like pose and illumination. Cross-modal person re-identification provides approaches for comparing data from two different modalities in complex settings, but is not directly applicable to our behavioral data.

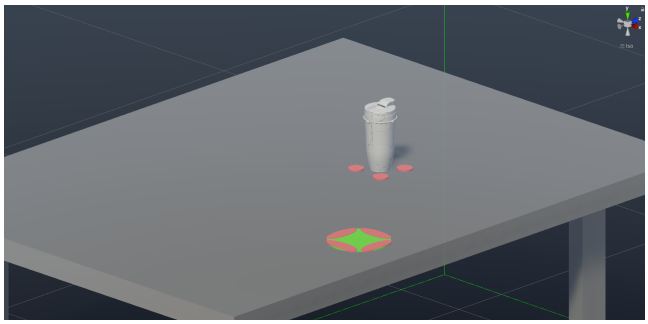

Figure 2: The VR environment with the water bottle placed at a starting position.

3 DATASET

The dataset consists of recordings of subjects performing the task of picking up an object and moving it to a desired location. Figure 1 shows the six objects that were looked at for this task and their varying physical dimensions —a vase, a water bottle, a box of macaroni and cheese, an applesauce cup, an ice cube tray, and isopropyl alcohol bottle. Figure 3 shows the eight configurations of pick-up and put-down locations in the real-world environment.

3.1 VR environment

The virtual reality environment was constructed using Unity. In it an object, starting point and end point would appear on a 0.7 meter tall table. The objects to be moved are 3D scans of the real-world objects and are sized one-to-one with the real objects. Figure 2 shows the VR environment. The user is instructed to pick up the object and to leave it in a specific placement area. This configuration shares the same dimensions as the real-world environment. Once the user has released the object in the placement area, they can not pick it up again. The object disappears after a half second pause, and one second after the disappearance, a new object and configuration appears.

The position and orientation of the three VR devices, headset, left-hand controller and right-hand controller, are recorded at a rate of 90Hz. Recording starts when an object and configuration are spawned and stops when the object disappeared. The moment the user picked up the object and successfully placed the object are also recorded to facilitate looking at different parts of the interaction, approaching pick-up, moving the object and post-placement.

3.2 Real environment

The real environment consists of a table with tape marking eight configurations. On a laptop, the subject is informed of the object and the task configuration. Once the object is placed on the starting location, the subject raises their hands above their head and back to their sides before performing the task. Once the object has been moved the subject again raises their hands above their head. The raising of hands is to ease the burden of manually marking interaction start for alignment with the VR data by providing easily detectable joint keypoints.

The real-data was recorded using two Azure kinects which recorded color and depth images as well as providing body tracking information and segmentation masks for humans in frame. The depth images are aligned with the RGB images using the API provided with the kinects, so the pixels have a one-to-one pixel correspondence between the two modalities. From the depth and color images, point clouds can be reconstructed and the background can be removed using the segmentation mask.

The cameras were positioned insert positioning information and were calibrated using a checkerboard based on the RGB cameras.

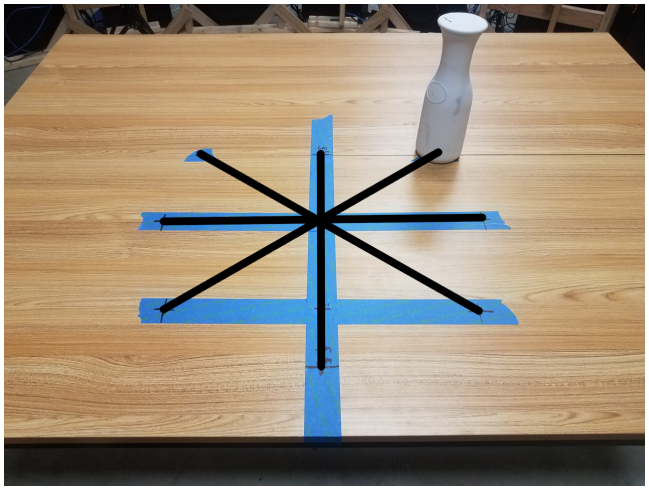

Figure 3: The real-world environment with the vase placed at a starting position. All potential configurations are marked with lines where the starting and stopping positions can be either ends of each line.

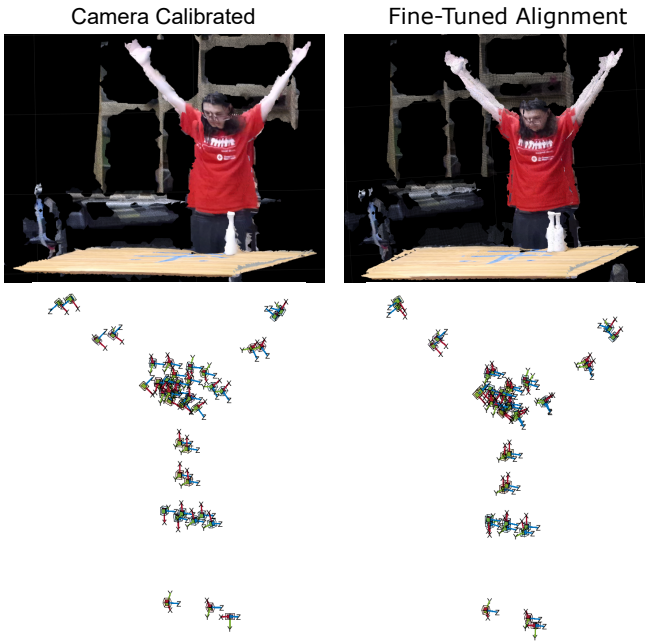

Figure 4: Alignment of the pointclouds and the tracked joints from the two Azure Kinect views.

This is used to get a rough alignment of the pointclouds and body joints from the individual kinects. A fine-tune alignment is obtained by performing the iterative closest points algorithm between the joints which the kinect reported as being properly tracked. Figure 4 shows an example of the initial pointcloud alignment and fine-tuned alignment. In this case, the fine-tuned pointcloud appears to have a poorer alignment, most noticeable in the tape on the table, while the fine-tuned skeleton appears better aligned. This is due to the bones from which the fine-tine transform is obtained being sparse at the waist and spine and denser at upper body.

The data recorded from the kinects are temporally aligned as both are able to be run using a single computer with a GTX 1080 TI GPU running the body tracking algorithm. Interactions will be manually trimmed to get similar formatting as the VR data with the start frame being the subject in a neutral pose after raising their hands above their head, the object movement being marked as the frame where the subject grabs the object, the task finish being when the subject releases the object at the stop position, and the interaction ending being the subject returning to a neutral pose after the object has been placed.

3.3 Capture Protocol

Here we describe the capture protocol used for this experiment. We recorded one session of VR data using two subjects and one session of real-world data using one subject. For the VR data, each subject performed four trials for each object-configuration combination for a total of 192 recordings (6 objects ×8 configurations ×4 trials). The order was generated by shuffling a list of all the combinations a subject performed. The subject performed all interactions back-to-back with no breaks.

For the real-world data, the subject performed two trials per object-configuration combination for a total of 96 trials (6 objects ×8 configurations ×2 trials). The combination order was again determined by shuffling a list with all combinations. Between each trial the subject set up the environment and then raised their hands above their head and returned them to be by their waist. Occasionally longer breaks were taken between trials for the subject to check that the video capture was continuing without issues.

3.4 Proposed changes to the environment

Here we describe changes to be made to the environment for further data collection.

3.4.1 Task Setup

One modification to the task set up would be to make the task a bit more involved. This could be done by creating an L shape with the tables so the subject will have to turn to place objects or switch which hand is holding the object.

3.4.2 VR Environment

Before a larger scale data capture the VR environment should be updated to be a bit more friendly to subjects. This can be done by making sure that the objects are inert when they spawn, currently they will move about a bit and the taller objects like the vase can fall over. Additionally, the subjects should be presented with a UI displaying their progress. The entire VR capture should also be broken up with breaks for each subject within a single session.

3.4.3 Real-World

One update to the real-world setting would be having a fuller view of the subject performing the task. In the current environment the table occludes the subject below the waist, and for smaller subjects would even lead to hand occlusion in the resting pose. One way of accounting for this would be to have a larger angle between the two viewpoints of the kinects. Another option would be increasing the height the cameras are placed at and having them look down at a steeper angle. For both of these options, the kinects tracking capabilities should be verified to ensure that the subjects joints are accurately marked even at these more extreme viewpoints. Adding a third kinect could also be explored although an additional kinect might require expanding to using an additional computer which would make temporal synchronization more complicated.

3.5 Proposed capture protocol

The data capture will be done over five sessions. The first two consists of only VR interactions or only real-world interactions and were separated by at least a week of time. The order of modality presented to the subjects was chosen randomly on a per-subject basis. This was to study how individuals interacted in each setting while limiting the influence of interactions with the other modality. The last two sessions consisted of interactions in both VR and the real-world. Here subjects performed a pick up and placement in one modality and immediately perform the task for the same object and configuration in the other modality. This will allow us to see how much influence recent performances of the task have on interactions across the modalities. For one of these sessions the real-world interactions will be presented first and for the other the VR interactions are first. This order is chosen randomly.

Qualitative data for these four sessions are captured by asking subjects how similar they thought the VR simulation was to the real-world after every VR interaction using a five-point Likert scale. The questions asked are:

1. The motion I made was similar to the motion for real-world interactions
2. The overall VR interaction is similar to real-world interaction
3. I feel like I'm performing the task in the real-world

The final session consists of VR interactions only. For this a ghost interaction will presented to the subject and they were asked to follow the motion the ghost made. Ghost interactions will be created by transforming previously recorded real-world interactions or by using previously recorded VR interaction data. Here subjects are asked whether the ghost was created from the real-world interactions or VR interactions and how comfortable it was.

The dataset consists of N subjects. For the first two sessions, the subject performed $\#Tasks$ repetitions of the task with a break every $\#BreakEveryX$ tasks. The multi-modality sessions consist of $\#MMSetReps$ task completions, where the $\#MMSetReps/2$ are in modality A and the other half are in modality B. The tracing session has subjects perform the task $\#Traces$ times.

4 Proposed Section - User Perception of Real-world interactions and VR interactions

Here we would describe the results we get from user perception on the tasks.

5 Real-world interactions compared with VR interactions

To compare real-world interactions and VR-based interactions we utilize two subjects data where subject A has provided data in VR and the real-world using their right-hand and where subject B has provided VR data using their right-hand. Here we are looking to quantify how consistently a subject performs a task and how comparable interactions in the different modalities are.

5.1 Methodology

We look to see how consistent a subject moves objects for fixed configurations. We look at user consistency within VR interactions, within real-world interactions and across the modalities.

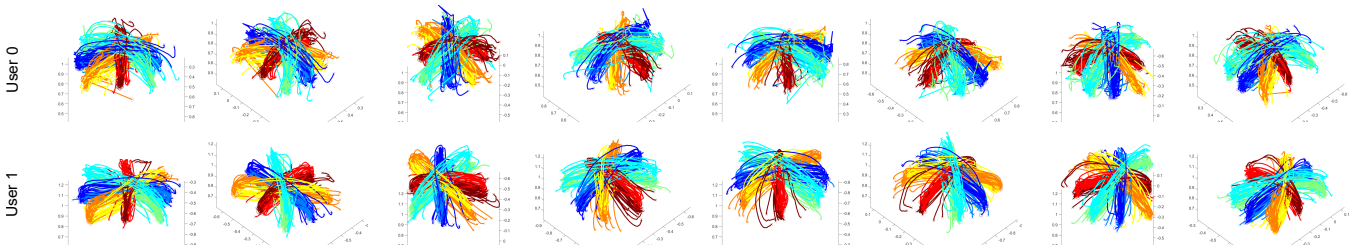

Figure 5: Right hand motions of subjects moving objects. The coloring varies with configurations across all objects and trials.

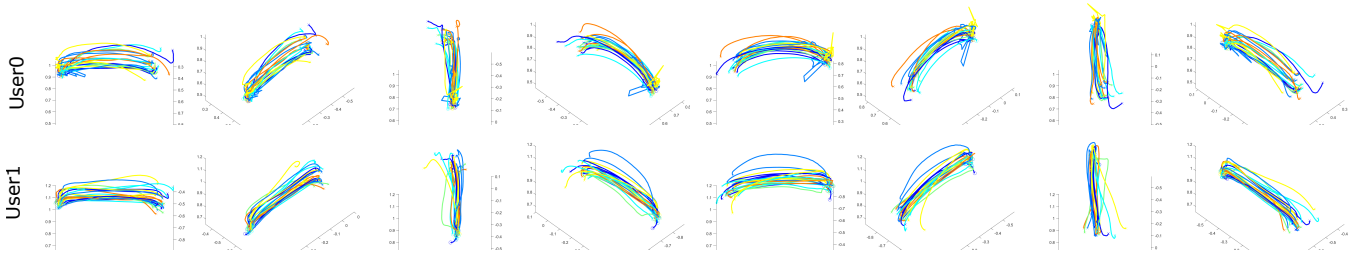

Figure 6: Right hand motions of subjects moving objects for one configuration. The coloring varies with the object being moved across the four trials.

When performing within-modality comparisons, we use the data as is. We look at consistency using dynamic time warping (DTW) for the positional values and something for the orientations. Here we look only at subsections of the recording where the subject is actively moving the object and using only the right hand controller data of each subject. For cross-modality comparisons we obtain a non-scaled transformation by performing iterative closest points (ICP) between the two data samples and then obtaining an error using DTW between the position of the right-hand VR controller and the right-hand landmark from the Kinect.

	User 1		User 2	
	Same Obj	Diff Object	Same Obj	Diff Obj
Config 1	0.23 ± 0.20	0.30 ± 0.15	0.12 ± 0.08	0.16 ± 0.05
Config 2	0.31 ± 0.35	0.40 ± 0.34	0.14 ± 0.13	0.19 ± 0.11
Config 3	0.26 ± 0.21	0.32 ± 0.17	0.12 ± 0.09	0.16 ± 0.06
Config 4	0.24 ± 0.18	0.32 ± 0.15	0.09 ± 0.07	0.13 ± 0.05
Config 5	0.18 ± 0.15	0.25 ± 0.11	0.14 ± 0.10	0.18 ± 0.07
Config 6	0.21 ± 0.17	0.29 ± 0.15	0.12 ± 0.10	0.17 ± 0.09
Config 7	0.21 ± 0.16	0.27 ± 0.12	0.13 ± 0.10	0.17 ± 0.06
Config 8	0.26 ± 0.20	0.33 ± 0.14	0.13 ± 0.10	0.16 ± 0.07

Table 1: Comparison of error from DTW function for same and different objects being moved in the same configuration. This is computed using the right hand only for the portion of the interaction where the subject is moving the object.

5.1.1 Within Modality —VR Motion

Table 1 shows the distance between a same-subject same-configuration same-object interaction pairing and a same-subject same-configuration different-object interaction pairing. By comparing the distances for each user we can see have that on average User 1 sees an increase of 0.07 for different object interactions and User 2 sees an increase of 0.04. Additionally, we can observe that certain configurations lead to more divergent behavior. For Users 1 and 2, the largest distance comes from Configuration 2. This is where the user moves the object from right to left at a constant distance from the table edge.

Table 2 has the same format as Table 1 but in this case the DTW function excludes the first and last 10% of the subjects motion. This

	User 1		User 2	
	Same Obj	Diff Object	Same Obj	Diff Obj
Config 1	0.05 ± 0.04	0.06 ± 0.03	0.05 ± 0.03	0.06 ± 0.03
Config 2	0.20 ± 0.40	0.25 ± 0.43	0.06 ± 0.08	0.08 ± 0.08
Config 3	0.06 ± 0.05	0.09 ± 0.04	0.06 ± 0.04	0.07 ± 0.03
Config 4	0.07 ± 0.09	0.10 ± 0.09	0.03 ± 0.02	0.05 ± 0.02
Config 5	0.06 ± 0.07	0.08 ± 0.07	0.05 ± 0.04	0.06 ± 0.03
Config 6	0.07 ± 0.12	0.09 ± 0.13	0.04 ± 0.03	0.06 ± 0.02
Config 7	0.05 ± 0.04	0.07 ± 0.03	0.04 ± 0.03	0.05 ± 0.02
Config 8	0.05 ± 0.04	0.06 ± 0.02	0.05 ± 0.04	0.06 ± 0.03

Table 2: Comparison of error from DTW function for same and different objects being moved in the same configuration with the first and last 10% of the motion removed. This is computed using the right hand only for the portion of the interaction where the subject is moving the object.

was done to see whether the grabbing and releasing positions contributed to the variance between the same object and different object comparisons. In this case we see an increase of 0.02 for User 1 and 0.02 for User 2. This indicates that while there is variability for the movement from the starting point to stopping point, a large source of divergence in behavior comes from the ends of the interaction. These tables show that user's have variations in how they perform the movement of different objects in VR. One area for further exploration would be to see if more complex patterns would decrease the difference users exhibit when moving unique objects.

5.1.2 Within Modality —Real Motion

To be filled in when manual marking of real-data is complete.

5.1.3 Cross Modality

To be filled in when manual marking of real-data is complete.

REFERENCES

[1] S. Bhalla, M. Goel, and R. Khurana. Imu2doppler: Cross-modal domain adaptation for doppler-based activity recognition using imu data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–20, 2021.

[2] S. Choi, S. Lee, Y. Kim, T. Kim, and C. Kim. Hi-cmd: Hierarchical cross-modality disentanglement for visible-infrared person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10257–10266, 2020.

[3] R. Delgado-Escano, F. M. Castro, N. Guil, V. Kalogeiton, and M. J. Marín-Jiménez. Multimodal gait recognition under missing modalities. In *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 3003–3007. IEEE, 2021.

[4] Y. Hao, N. Wang, X. Gao, J. Li, and X. Wang. Dual-alignment feature embedding for cross-modality person re-identification. In *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 57–65, 2019.

[5] T. Hoang, T. Nguyen, C. Luong, S. Do, and D. Choi. Adaptive cross-device gait recognition using a mobile accelerometer. *Journal of Information Processing Systems*, 9(2):333–348, 2013.

[6] S. Li, A. Ashok, Y. Zhang, C. Xu, J. Lindqvist, and M. Gruteser. Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–9. IEEE, 2016.

[7] J. Liebers, M. Abdelaziz, L. Mecke, A. Saad, J. Auda, U. Gruenefeld, F. Alt, and S. Schneegass. Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2021.

[8] R. Miller, N. K. Banerjee, and S. Banerjee. Using siamese neural networks to perform cross-system behavioral authentication in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 140–149. IEEE, 2021.

[9] T. Mustafa, R. Matovu, A. Serwadda, and N. Muirhead. Unsure how to authenticate on your vr headset? come on, use your head! In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, pp. 23–30, 2018.

[10] I. Olade, C. Fleming, and H.-N. Liang. Biomove: Biometric user identification from human kinesiological movements for virtual reality systems. *Sensors*, 20(10):2944, 2020.

[11] E. Pezent, A. Israr, M. Samad, S. Robinson, P. Agarwal, H. Benko, and N. Colonnese. Tasbi: Multisensory squeeze and vibrotactile wrist haptics for augmented and virtual reality. In *2019 IEEE World Haptics Conference (WHC)*, pp. 1–6. IEEE, 2019.

[12] C. E. Rogers, A. W. Witt, A. D. Solomon, and K. K. Venkatasubramanian. An approach for user identification for head-mounted displays. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pp. 143–146, 2015.

[13] A. Sepas-Moghaddam and A. Etemad. Deep gait recognition: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[14] G. Wang, T. Zhang, J. Cheng, S. Liu, Y. Yang, and Z. Hou. Rgb-infrared cross-modality person re-identification via joint pixel and feature alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3623–3632, 2019.

[15] A. Zenner and A. Krüger. Drag: on: A virtual reality controller providing haptic feedback based on drag and weight shift. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.

[16] A. Zenner, K. Ullmann, and A. Krüger. Combining dynamic passive haptics and haptic retargeting for enhanced haptic feedback in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2627–2637, 2021.