Understanding User Identification in Virtual Reality Through Behavioral Biometrics and the Effect of Body Normalization

Jonathan Liebers
University of Duisburg-Essen
Essen, Germany
jonathan.liebers@uni-due.de

Mark Abdelaziz
German University in Cairo
Cairo, Egypt
mark.abdelaziz@student.guc.edu.eg

Lukas Mecke
Bundeswehr University Munich
Munich, Germany
lukas.mecke@unibw.de

Alia Saad
University of Duisburg-Essen
Essen, Germany
alia.saad@uni-due.de

Jonas Auda
University of Duisburg-Essen
Essen, Germany
jonas.auda@uni-due.de

Uwe Grünefeld
University of Duisburg-Essen
Essen, Germany
uwe.gruenefeld@uni-due.de

Florian Alt
Bundeswehr University Munich
Munich, Germany
florian.alt@unibw.de

Stefan Schneegass
University of Duisburg-Essen
Essen, Germany
stefan.schneegass@uni-due.de

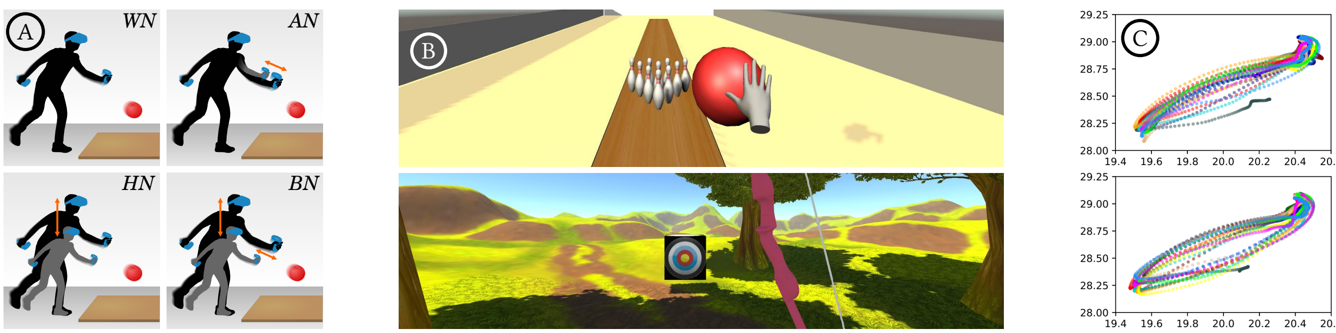


Figure 1: We investigate the identification of users performing tasks in VR. (A) We normalize their virtual body proportions (WN: Without Normalization, AN: Arm length Normalization, HN: Height Normalization, BN: Both Normalizations, black: real body proportions, gray: applied normalization). (B) Tasks consist of a Bowling and an Archery scenario. (C) The captured spatial motion data enables implicit identification (upper: model training data, lower: validation data from another day).

ABSTRACT

Virtual Reality (VR) is becoming increasingly popular both in the entertainment and professional domains. Behavioral biometrics have recently been investigated as a means to continuously and implicitly identify users in VR. Applications in VR can specifically benefit from this, for example, to adapt virtual environments and user interfaces as well as to authenticate users. In this work, we conduct a lab study ($N = 16$) to explore how accurately users can be identified during two task-driven scenarios based on their spatial movement. We show that an identification accuracy of up to 90% is possible across sessions recorded on different days. Moreover, we investigate the role of users' physiology in behavioral biometrics by virtually altering and normalizing their body proportions. We find that body normalization in general increases the identification rate, in some cases by up to 38%; hence, it improves the performance of identification systems.

CCS CONCEPTS

• **Human-centered computing** → *Interaction techniques*; *Virtual reality*; • **Security and privacy** → *Usability in security and privacy*.

KEYWORDS

identification, virtual reality, usable security, task-driven biometrics

ACM Reference Format:
Jonathan Liebers, Mark Abdelaziz, Lukas Mecke, Alia Saad, Jonas Auda, Uwe Grünefeld, Florian Alt, and Stefan Schneegass. 2021. Understanding User Identification in Virtual Reality Through Behavioral Biometrics and the Effect of Body Normalization. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3411764.3445528

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445528

1 INTRODUCTION

Virtual Reality (VR) has undergone a substantial evolution in recent years and is increasingly becoming a part of users' daily lives. Although games are currently the main driver for VR headsets, other applications are also gaining importance. Examples range from training [37] to rehabilitation and therapy [34] to e-commerce [30].

As is the case for other technologies, VR can benefit from knowledge about the current user's identity, particularly for multi-user scenarios. Such scenarios could include work spaces in which VR headsets are shared or situations at home in which multiple family members or friends use VR headsets together. In such settings, knowing a user's identity creates opportunities for adapting the user interface to the user's needs, loading and setting personal preferences, or granting access to personal information (e.g., social media, personal messages, or financial information).

Current approaches to identify (or authenticate) users employ forms like PINs or passwords entered through hand-held controllers and a virtual keyboard. These methods can interrupt interaction with the system (e.g., through a password prompt popping up) and hamper user immersion in a virtual environment. Moreover, hand-held controllers as an input modality are readily observable by a bystander, making them inherently insecure [10].

These traditional approaches are generally implemented by making the user perform an explicit interaction with the identification mechanism, such as selecting their user name from a list or the entry of a password through a virtual keyboard. Nevertheless, the utilization of an implicit interaction for identification is more favorable, as it does not interfere with the user's interaction but can derive the necessary information for identification from the user's general interaction with an application.

In this work, we create an identification system, capable of implicitly identifying users by their behavior in VR. Specifically, we look into two different task-based scenarios. Our scenarios mimic common VR games, such as Bowling and Archery, where users naturally interact with the game (cf., Figure 1). Our identification system employs the elicited information from the user's interaction to implicitly determine the user's identity in the background. We use a standard consumer-grade head-mounted display (HMD) in a lab study ($N = 16$) and collect spatial motion data from both the HMD and the hand-held controllers. To understand the influence of physiology on behavioral biometrics in VR, we normalize each user's height and arm length so that all participants shared the same virtual body proportions.

We collected data in two separate study sessions to investigate identification performance across several days, and we achieved an accuracy of up to 90%. Also, we found that normalizing body proportions (precisely, normalizing height) for identification increases the identification accuracy by up to 38%.

Although this identification accuracy is not sufficient for security-critical use cases, the underlying principle allows for creating VR applications that derive the user's identity from their regular interaction with the application transparently in the background. Hence the burden of explicit identification is removed from the user, allowing developers of VR applications to create a painless process for the user and a seamless experience. Simultaneously, utilization of a body normalization can strongly enhance identification accuracy.

Contribution Statement. The contribution of this work is twofold. First, we create a prototype for implicit user identification in VR through task-driven behavioral biometrics in two scenarios and report on a user study ($N = 16$). Our study, which spread across two different days, includes investigating of how modified body proportions influence identification accuracy. Second, we provide guidelines and discuss how the interactive VR system's identification rate can be enhanced by employing body normalization. Moreover, we publish our elicited data set.

2 RELATED WORK

Our work is situated within the domain of behavioral biometrics for the purpose of identification in VR. As we elicit spatial motion data, we also look at task-driven biometrics.

2.1 Identification and Authentication in VR

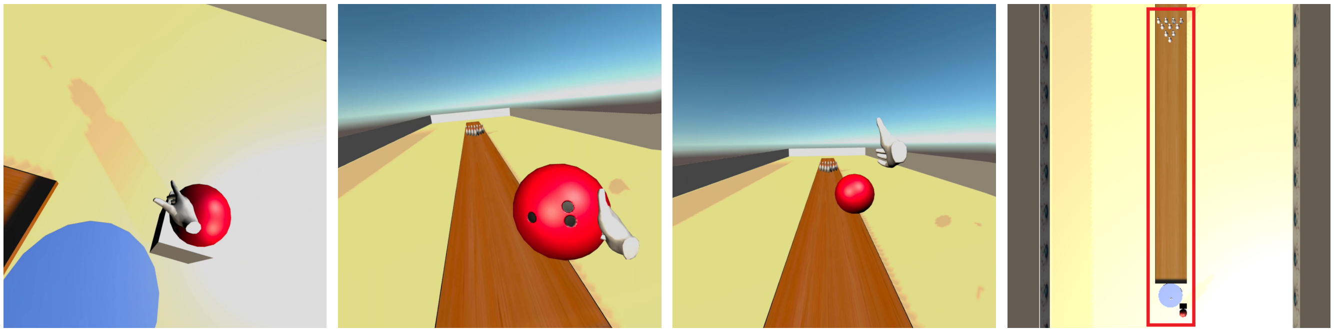
Authentication methods for VR are mostly adaptations from mobile devices, implementing knowledge-based schemes (e.g., PIN and password entry or pattern locks) [40]. Previous research has shown that traditional methods for authentication in VR, such as the entry of a PIN or password, are slow and subject to threats commonly known from the real world, such as shoulder surfing [10]. In particular, prior work showed that up to 18 % of authentication attempts can be observed from a shoulder-surfing bystander [10].

Virtual reality systems are in general characterized by a high degree of immersion [6]. Therefore, to keep this immersion, implicit authentication schemes that do not interrupt the interaction of the user [36] and are carried out through actions that the user would perform anyways seem to be a particularly good fit for VR [15]. Besides the benefits of implicit authentication, behavioral biometrics also enables continuous authentication [8].

The disadvantages of currently used authentication methods in VR and the need for implicit authentication methods motivates the development of new methods for identification and authentication in VR. Pfeuffer et al. have shown that body motion and proprioception can be used for authentication in VR by measuring spatial relations between the controllers and the HMD movement [31]. Similarly, Mustafa et al. developed a solution to derive identity from the head movement patterns which can be collected from the HMD [28]. Sivasamy, Sastry, and Gopalan developed a similar approach for the goal of continuous authentication [38]. These concepts have been extended to include knowledge-based components, such as 3D passwords [4] in conjunction with gaze and head pose [9]. Furthermore, Mathis et al. include hand motion during the input of a PIN in VR as a second modality [23].

2.2 Task-driven Biometrics

In recent years a new subcategory of behavioral biometrics moved into the focus of research. Task-driven biometrics are based on data elicited through a user's performance of a manual task. Igarashi et al., for instance, measured individual characteristics with regards to a driving task to identify the driver of a car [14]. Pohl et al. found that even a single button press contains a high degree of individual behavior [32]. Kupin et al. tasked users to throw a ball in VR and were able to recognize them across different sessions that took place days apart [20]. An extended analysis of this task



(a) (b) (c) (d)

Figure 2: Illustration of the Bowling task. Figure (a) depicts the first discrete phase that spans from the initiation of the task until the red ball on the pedestal is grabbed. The participant would then execute the swinging motion (b) and release the ball (c). The final phase forms the duration until the ball either hits the pins or leaves the alley by intersecting the red border (d).

and experimental setup was performed by Ajit et al. [3], yielding a higher accuracy. Miller et al. created an extended solution for rejecting intruders [26] and performed a within- and cross-system evaluation [27]. "BioMove" utilized the kinesiological movements of the user (captured through an HMD and controllers) while grabbing, moving and dropping balls and cubes from one position into another container [29]. Moreover, they included eye tracking in their system and recorded gaze (i.e., where users were looking during the task).

The behavior we classify always occurs with respect to a problem, or in case of task-driven biometrics, a task. Mecke et al. have shown that typing behavior can be intentionally altered to fit a given target behavior [25] and Prange et al. lead an investigation into the specifics of individual user behavior with respect to goal-oriented tasks [33]. In a broader sense, also mimicry attacks (i.e. mimicking a legitimate user) use task-driven behavior changes, though, the goal is to gain access to a biometric system rather than the change itself. Related work has shown such attacks to be successful for several systems, including touch input behavior [16] and keystroke dynamics on PCs [39] and mobile phones [17].

2.3 Virtual Embodiment

Kilteni et al. defined the sense of embodiment, stating that the sense of embodiment emerges when the body's properties are processed as if they were the properties of one's own biological body [18]. The virtual embodiment is dependent on the realism of the virtual representation of the real body part [5]. For example, VR users identify stronger with a realistic virtual hand than with a more abstract virtual representation. Such factors can play an important role in various fields that apply VR such as medicine [22].

Previous research also investigated the manipulation of virtual body parts. Kilteni et al. experimented with "a very long arm illusion", investigating the limits of changing the VR user's arm length up to the point that it is not experienced as one's own anymore [19]. Rothe et al. found that a decreased eye height while watching 360° recordings in VR was less disturbing to the viewers than an increased one [35]. Viewers preferred that the virtual eye height matches their real eye-level. Furthermore, if that was not the case, a lower virtual eye height was preferred over a higher one.

Similar to these approaches, we modify the virtual arm length and the body height for the purpose of normalization across participants to investigate behavior based identification.

2.4 Summary

Prior work identified behavioral biometrics as a suitable means to identify users implicitly. Applying such behavioral biometric-based approaches to VR comes with many benefits: based on knowledge on the identity, adaptive VR interfaces can be built, and identification might be realized without breaking the immersion and with minimal effort for the user. This opportunity has been recognized by the community, as was shown by prior work, that the application of behavioral biometrics to VR is generally possible.

What is missing as of today is a more nuanced understanding of this approach, in particular, (a) how it performs beyond artificial tasks and (b) how the physiology of the user influences accuracy. To close this gap, we contribute an investigation of how well users can be identified during ecologically valid tasks. In addition, VR provides an unprecedented chance to investigate the influence of physiology since it allows physiological properties of users to be freely manipulated.

3 APPARATUS

This section describes how we designed an apparatus to investigate (a) realistic VR tasks by designing meaningful scenarios and (b) the influence of physiology on behavioral biometrics by normalizing physiological features in VR.

3.1 Scenarios

We implemented two different task-based scenarios for the users to solve : a Bowling and an Archery task. As an engine, we utilized Unity3D and targeted the Oculus Quest as a consumer-grade entry-level HMD. The scenarios we implemented were designed to be familiar and straightforward to solve . Bowling employs a coarse-grained movement and Archery, in turn, requires a precise, fine-grained movement to be completed successfully.

(a) (b) (c) (d)

Figure 3: Illustration of the Archery task. First the bow needs to be moved towards the quiver to pick up an arrow (a). The second phase spans from the mounting of the arrow at the bow until the string is being pulled (b). During the third phase, the string is pulled to tension the bow (c). The fourth phase spans from the release of the arrow until it hits a target (d).

3.1.1 Bowling-Task. In the bowling task (cf., Figure 2), users spawn in a virtual bowling alley with 1.3 m width and a length of 7 m between the start and pins. The bowling ball has a virtual mass of 5 kg and an angular drag of 0.05. It spawns on the right side on a pedestal that has a height of 0.75 m. If participants miss a shot completely, the ball will despawn once it exceeds a boundary (cf., Figure 2(d)), otherwise hit pins are replaced for each new trial.

In this controller-based scenario, users have the goal to hit as many pins as possible. The task can be decomposed into three discrete phases (cf., Figure 2). In the first phase, the user grabs the ball with the controller by initiating a pressing motion on the grip button. The second phase consists of the swinging motion, while the grip button is being pressed until the ball is released. Once the ball is released, the third phase is initiated, which remains active until the ball either hits the pins or goes off the alley. After the third phase is complete, a new round is initiated.

3.1.2 Archery-Task. Archery (cf., Figure 3) is a controller-based task that requires high precision. Users are placed in a forest environment, approximately 3 m in front of a target. Their goal is to shoot an arrow at the bull's eye. The bow is mounted to the user's left hand while the right hand is free to pull the string. The arrows are attached to the bow by moving it towards a virtual quiver (located on the left-hand side, cf., Figure 3).

This task can be decomposed into four discrete phases. The first phase is active until the bow is moved towards the quiver, and an arrow attaches to the bow. The second phase ends when the user grabs the string of the bow with the right hand. The third phase describes the period of the string being pulled until it is released or the hand is moved too far away from the back of the bow. The task concludes with the fourth phase until the arrow hits an object.

3.2 Body Normalization

We impose two different body normalization types to explore the influence of physiology in behavioral biometrics: 1) Arm length Normalization ("AN") and 2) Height Normalization ("HN"). For applying the normalizations, our system acquires physiological data by measuring users' height and arm length through the HMD. Therefore, the application requests users to stand up straight and stretch their arms in front of them for 10 seconds. Consequently, the system determines their heights as reported by the HMD and measures the maximum distance from controllers to HMD.

3.2.1 Arm length Normalization. The first type of normalization is the normalization of the user's virtual arm length. The normalization is achieved by modifying the positional mapping of the virtual hand onto the real-world controller. The effect is that the virtual maximum arm extension length is the same for all users. This is established through a linear change in the acceleration of the user's hands up to the point that the arm is fully extended. Here, the coordinates of the left and right hand of the user are calculated by the formula $P_{Hand} = P_{Head} + d \cdot D \cdot F$, where P_{Hand} represents the position of the hand, P_{Head} is the position of the head (i.e., the spatial coordinates of the HMD), d is a direction vector pointing in the direction of the hand, D represents the previously measured maximum distance from the head to the hand and F is an automatically calculated normalization factor. Subsequently, this normalization allows the assignment of a consistent virtual arm length to all users. Thereby, a short-armed person receives longer virtual arms, whereas a long-armed person receives shorter virtual arms. Although not primarily intended, this normalization also assures that every person has the same virtual capability to solve the given tasks, as the required arm length is normalized independent of the real capabilities.

3.2.2 Height Normalization. The second type of normalization is the normalization of the user's body height. This normalization allows the assignment of the same height across all users. The modified reference point is the virtual head of the user. In contrast to the Arm length Normalization, another design approach is required for the implementation of this normalization because changes to the positional mapping (e.g., by increasing the acceleration) between the virtual and real head can induce unwanted side effects such as cybersickness. To mitigate these effects, we implemented this normalization by setting the tracking origin type of the "OVR-CameraRig" object within Unity3D to 'eye level' and assigning its y-position a constant value. This results in an unhindered change of position of the virtual head up to the constantly set virtual height but may result in the impression that the virtual feet are below the level of the floor (i.e., a user that is taller than the set height would be able to reach past the virtual floor by bowing down). To counter this drawback, the tasks were designed so that the users neither have to perform this motion nor do they have to pick something up from the level of the real floor.

4 STUDY

We verified our approach to task-driven biometrics and explored the effects of the changes imposed by the body normalization.

4.1 Study Design

To verify our approach of task-driven biometrics, we conducted a within-subject controlled laboratory study in Virtual Reality using the Oculus Quest. Our study followed a repeated-measures design and was split into two sessions that took place on two different days, sharing the same study design in both sessions. All participants took part in both sessions. We chose this split to prove the stability of our approach and to gain realistic data as the behavior might change between days. This allows us to explore whether we would be able to re-identify users across separate days.

Our independent variables were *Scenario* with two levels (*Bowling* vs. *Archery*) and *Type of Normalization* with four levels (*Without Normalization* vs. *Arm length Normalization* vs. *Height Normalization* vs. *Both Normalizations*). Each *Scenario* was tested in a block. All blocks were counterbalanced using a Latin square. In each block, we tested all *Types of Normalization*, again using a Latin square design. Here, we took all combinations of both Latin square designs, resulting in eight configurations.

Our research questions were:

RQ1 To what extend are different Virtual Reality scenarios feasible to implicitly identify users?

RQ2 To what extend do physiological factors influence Behavioral Biometrics?

4.2 Study Setting

We chose the Oculus Quest head-mounted display (HMD) as target device for Virtual Reality. It features two six Degrees-of-Freedom (DoF) controllers, supporting orientation and positional tracking based on an inside-out tracking system and offering a field-of-view of 100° at a refresh rate of 72 Hz. It operates without being connected to a computer, thus being a wireless, consumer-grade device.

The study took place in a room with 3m×3m of free space. The experimenter prepared the Oculus Quest for the procedure by enabling the integrated screen recording of the virtual environment and covering the internal proximity sensor so that the device does not switch into standby mode. Furthermore, the experimenter set the safety guard beforehand and took care that the participants would not leave the designated area during the study.

4.3 Procedure

Before the beginning of the actual study, participants gave written and informed consent. We specifically informed the participants that they could cancel their participation in the study at any time without detriments. Also, we answered any question from the participants with regards to the procedure but we did not tell them about the imposed types of body normalization prior to the study. Moreover we informed the participants that the virtual environment and their actions in the virtual environment would be recorded, both data-wise and video-wise (external and in-app recording) . After a short introduction to the Virtual Reality headset, assisted adjustment of the straps and the adjustment of the device itself (i.e., the inter-pupillary distance), we tested two blocks, one for each scenario (*Bowling* and *Archery*). After each block, the participants took off the headset and filled out a Raw NASA TLX questionnaire [12], rating each scenario's workload. The Raw NASA TLX is a commonly used modification of the NASA TLX and the employed scale was the standardized 20 pt. scale [11].

Participants could always try out the scenario first. Afterwards, we tested four types of normalization (i.e., *Without Normalization*, *Height Normalization*, *Arm length Normalization*, and *Both Normalizations*) for 12 trials each. We counterbalanced the order of the blocks and the order of the types of normalization within each block using a Latin-square design. After participants had finished all blocks, we conducted a semi-structured interview. We repeated the procedure in a second session without the questionnaires and interviews. Each participant took approximately 45 minutes to finish the first session and about 30 minutes for the second session.

4.4 Participants

We recruited 16 volunteers (2 female, 14 male) via University mailing list (mean age=25.47, SD=3.44, all were right-handed). We asked participants for their height as reported in their personal ID card, which yielded an average value of 177.85 cm (SD=8.23 cm). On a scale from 1 (low; never experienced VR before) to 10 (high; daily experience of VR) they rated their previous VR experience at an average of 3.78 (SD=3.19). In total six participants responded that they had a form of visual impairment and one did report partial color blindness. During the study, five out of these six participants wore a visual aid (e.g., glasses or contact lenses) as a compensation. One participant did not wear any corrective eye wear but assured that this would not affect their performance in VR.

For the body normalizations, we applied 168 cm as the normalized height, which equals to the average human height in Europe subtracted by a measured margin for the Oculus Quest device [24] . Moreover, we set 0.7 m as the normalized arm length. We obtained this value through measurement, following the principle of Da Vinci's Vitruvian Man.

4.5 Ethics

To preserve participants' privacy, we assigned pseudonyms to the data of each participant at the time of elicitation. After finishing the study, we deleted the mapping of participants' true identity to the given pseudonyms so that no backtracking of participants' true identities is possible. Note, that one purpose of the system is identification. If employed in practice, it would be important to inform users about the fact that the collected data is or can not only be used for interaction, but also to identify them. Moreover, VR applications in general can bear a risk to privacy, as discussed by Adams et al., which needs to be accounted for [2].

5 ANALYSIS

We create a deep learning classifier that is able to predict the user's identity across the two sessions of our study. Furthermore, we evaluate its validation accuracy that is changed by the imposed body normalizations.

5.1 Data Set

For the following evaluation, we split our data set into four separate training and validation sets, one for each of the four types of body normalization[1]. For training classifiers, we always use the data from the first day of our study. Validation is exclusively performed through testing the data from the second day of the study. We opted for this split, as we assume that consecutive, repetitions of the task's execution are in general more likely to be similar than executions that were captured days apart. Furthermore, this strict split demonstrates the real-world applicability of our approach, as participants had to re-equip the HMD for the second day. Thus, the model cannot learn the specifics of one session such as wearing the HMD in an odd way. Each repetition in the study forms one sample in our data set. We always compare data within one type of normalization and never across conditions. In contrast to our split, a cross-validation with a standard split (e.g., 80:20) would lead to a mixing of data that was elicited on different days resulting in a model possibly using session dependent information. Most likely this would enhance the overall achieved results but would give the classifier an unrealistic advantage.

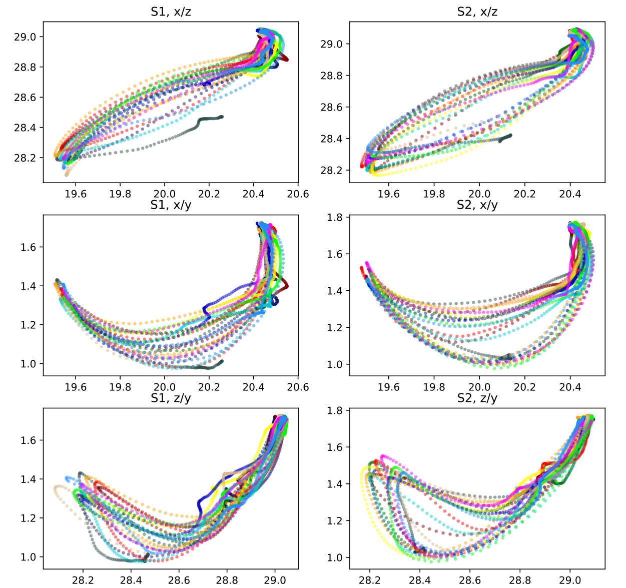
We then plotted all elicited data and performed a visual inspection. We did this in order to find random outliers that occured due to a tracking loss and removed them from the data set. Figure 4 visualizes an excerpt of the spatial data that we captured from two different participants and Figure 5 depicts the deviation of user height in the Archery scenario.

5.2 Preprocessing

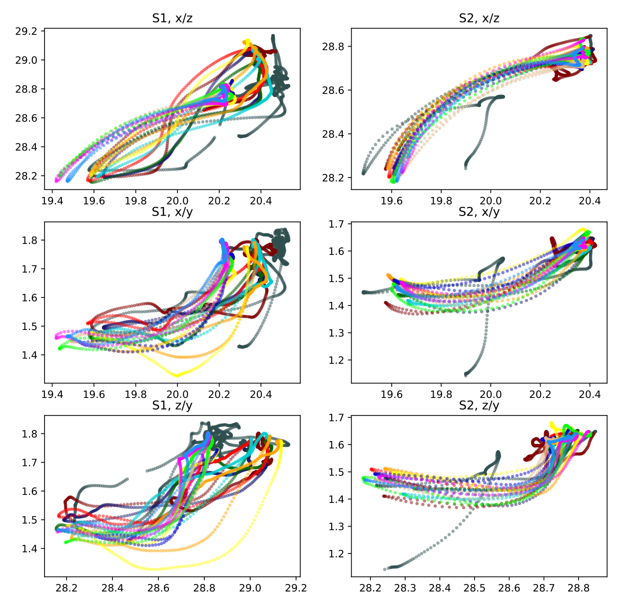
We apply the same preprocessing to all of the elicited data. First of all, we deduct the global coordinates of the hand-held controllers from the HMD to obtain their local coordinates with relation to the HMD. This means that the obtained data is invariant of its global position and that we do not classify users based on their absolute position within the tracking space. Moreover, we create several feature sets (cf., Table 1) to train the model on different parts of the available data. If we include the HMD in a featureset, we transform its coordinates with respect to its origin in the virtual environment, i.e., we subtract the global coordinates of the initial point of appearance from all subsequently captured points. This leads to the HMD being represented by its change in space over time (e.g., the HMD moved 30 cm in a certain direction) instead of its consecutive global coordinates. We furthermore normalize the Euler angle values in an interval of $[0, 1)$. This way, we reach positional invariance of the HMD so that only the motion that the participants apply is transferred to the model.

Each repetition in the study corresponds to one sample in the data set. As each repetition bears a different length and some repetitions form very large outliers, this consequently means that the shape of the data is hard to unify. We tried the common approach

[1] Our data set is publicly available. It can be retrieved from https://research.hcigroup.de.



(a) Participant 2



(b) Participant 11

Figure 4: Motion data of the left hand-held controller in the Archery scenario without any applied body normalization for participants 2 and 11 . The axes indicate the global coordinate frame. The colors indicate the repetitions within the study. "S1" and "S2" correspond to the first and second day of the study and "x/z", "x/y" and "z/y" are the plotted Unity axes (i.e., top view, side view, frontal view). The similarity within each participant is well visible together with the dissimilarity between participants.

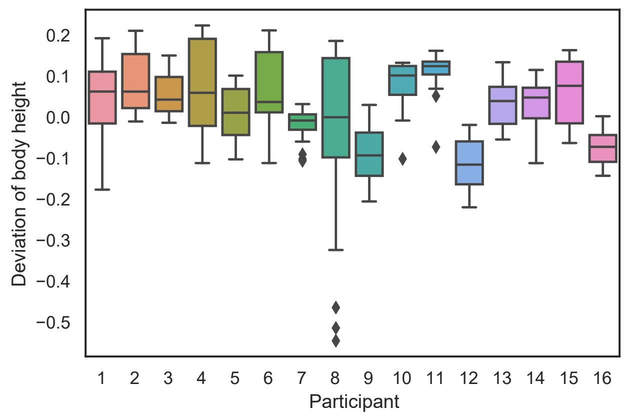


Figure 5: Deviation of user height for the Archery scenario across all repetitions.

of prepadding the data with zeros, yet a single outlier results in a large amount of padding for all samples. Moreover, this padding might have results on the training and validation, since the model then could learn to classify the length of the execution of a motion. We therefore changed our approach to reach temporal invariance.

To classify the fine grained motion of the participants , we implemented a window slicing approach [21]. Here we employ a rolling window with a size of 10 and a step size of 1 that iterates over each participant's data recording to generate new samples for each step. We apply the same preprocessing and the same sliding window to the validation data. As we classify each sample as generated by the sliding window, we furthermore perform a majority vote of the predicted labels to determine the user's identity.

5.3 Model Architecture

To classify the elicited data from the study we developed two deep learning models in Python, utilizing Keras and Tensorflow [1, 7]. Our first model consists of three long-short term memory layers ("LSTM") with 100 units each and is a recurrent neural network ("RNN") [13] . The first layer returns its sequences to the second layer and the second layer repeats this by returning its sequences to the third layer. All LSTM layers utilize the default sigmoid activation function. At last, a fully-connected layer utilizing the "softmax" activation function is reached that consists of 16 units, where each unit corresponds to the identity of one person. We used Adam as an optimizer with a learning rate of 0.001 and trained the model for 500 epochs.

The second model employs a multilayer perceptron ("MLP") and consists of a Flatten input layer, followed by three Dense layers that respectively consist of 256, 64 and 16 units. The first two dense layers use the rectified linear activation function ("ReLU") whilst the last utilizes "softmax" as an activation function. For this model, stochastic gradient descent ("SGD") is used as an optimizer with a learning rate of 0.001. As this model converges in general faster, we trained it only for 100 epochs. We set a constant random generator seed for the training of all model instances.

Table 1: Table of all evaluated feature sets with their descriptions and cardinalities (Card.). The HMD and both controller objects consist of three coordinates each (x, y, z) for their position and rotation.

Id.	Card.	Controllers	HMD	Timestamp	Phase
F0	19	×	×		×
F1	18	×	×		
F2	19	×	×	×	
F3	12	×			

5.4 Feature Sets

We created multiple groups of features to train the classifier on. Each group formed one set of features. Thus, we seek the optimal set of information the classifier requires for a high identification rate (cf. Table 1).

F0 The feature set F0 consists of the phase, the HMD and the controllers. Both the local position (x, y and z) and rotation ($euler\_x$, $euler\_y$ and $euler\_z$) of the HMD and the hand objects are part of this set. The phase describes the stage of the interaction for the Bowling and Archery scenario.

F1 The next feature set, F1, is a reduction of F0, as it takes the same features but removes the phase.

F2 The feature set F2 employs the same information as F0 but replaces the phase with a timestamp that denotes the passed time from the beginning of the interaction. The timestamp is intended as a hint for the order of the slices that are created from the rolling window so that relationships across slices can be learned by the neural network. The timestamp is a normalized number within the interval $[0, 1)$.

F3 Finally, the feature set F3 only consists of the vectors from the controllers towards the HMD.

6 RESULTS

First, we present our identification results that we obtained from our validated classifier. Next, we discuss the outcome of the semi-structured interviews and the Raw NASA TLX.

6.1 Identification Results

We validated our two model architectures with all feature sets (cf., Table 1). Table 2 refers to our achieved results. The highest overall identification accuracy is 0.90 for the Archery scenario with feature set F3 and the application of *Height Normalization* with the recurrent model (cf., Figure 6). For Bowling, the highest accuracy is 0.68 with feature set F2, where the Height Normalization was applied in combination with the recurrent model.

We applied inferential statistics to prove that the imposed body normalizations have an effect on the identification rate of the classifier. We apply the statistics to the per-participant identification rate, i.e., the diagonal axis in each obtained confusion matrix for all scenarios and feature sets. We first perform a Friedman test for each row in Table 2, once for the multilayer perceptron model and once for the recurrent model. If the Friedman test returns significant results ($p < 0.05$), we conduct six Wilcoxon tests for pairwise

Table 2: Overview of all validation accuracies (Acc.) for all scenarios, feature sets (F.-Set) and body normalizations "Without Normalization" (WN), "Arm Length Normalization" (AN), "Height Normalization" (HN) and "Both Normalizations" (BN). We differentiate between recurrent model "RNN" and multilayer perceptron "MLP". Highest accuracy per model is marked bold.

		Acc. WN		Acc. AN		Acc. HN		Acc. BN	
Scenario	**F.-Set**	MLP	RNN	MLP	RNN	MLP	RNN	MLP	RNN
Archery	F0	0.32	0.48	0.56	0.67	0.57	0.63	**0.68**	**0.86**
Bowling	F0	**0.55**	**0.65**	0.41	0.53	**0.55**	0.59	0.46	0.58
Archery	F1	0.38	0.54	0.56	0.63	**0.65**	0.69	0.64	**0.84**
Bowling	F1	0.49	0.59	0.42	0.50	**0.55**	**0.62**	0.48	0.60
Archery	F2	0.37	0.56	0.56	0.65	**0.66**	0.66	0.63	**0.86**
Bowling	F2	0.52	0.58	0.42	0.47	**0.54**	**0.68**	0.45	0.56
Archery	F3	0.61	0.63	0.68	0.70	**0.78**	**0.90**	0.68	0.84
Bowling	F3	0.55	0.56	0.43	0.47	**0.63**	0.66	0.55	**0.67**

Table 3: Overview of all significant Wilcoxon tests to describe the effect of the body normalizations. WN = Without Normalization, HN = Height Normalization, BN = Both Normalizations, RNN = Recurrent Neuronal Network (first model architecture), MLP = Multilayer Perceptron (second model architecture).

Scenario	**Model**	**F-Set.**	**Comparison**	**W**	**Z**	**p**	**r**
Archery	RNN	F0	WN vs. BN	8	-2.650	.034	.166
Archery	MLP+RNN	F0	WN vs. BN	32	-3.444	.002	.108
Archery	MLP+RNN	F0	HN vs. BN	40	-2.603	.048	.081
Archery	MLP+RNN	F1	WN vs. BN	58	-2.793	.025	.087
Archery	MLP+RNN	F1	WN vs. BN	49	-2.795	.025	.087
Archery	MLP+RNN	F2	WN vs. BN	49	-2.795	.025	.087
Bowling	MLP+RNN	F3	AN vs. HN	55	-2.680	.037	.084

comparisons of all body normalizations. Due to the large amount of combinations, we only report significant results. Table 3 provides an overview.

A Friedman test for the recurrent model with feature set F0 for Archery showed a significant difference ($\chi^2(3) = 8.168, p = 0.043, N = 16$). The pairwise comparison of the four types of body normalization through Wilcoxon tests revealed one significant effect of an increased identification: *Without Normalization vs. Both Normalizations* ($W = 8, Z = -2.650, p < 0.034, r = 0.166$). To assess the effect that both models are subject to, we conducted another Friedman test on the fused confusion matrices of the MLP and recurrent model for Archery in F0. We found a significant difference ($\chi^2(3) = 10.951, p = 0.0120, N = 16$). Comparing *Without Normalization vs. Both Normalizations* through Wilcoxon tests lead to a significant result ($W = 32, Z = -3.444, p < 0.002, r = 0.108$). Comparing *Height Normalization vs. Both Normalizations* lead again to a significant difference ($W = 40, Z = -2.603, p < 0.048, r = 0.081$).

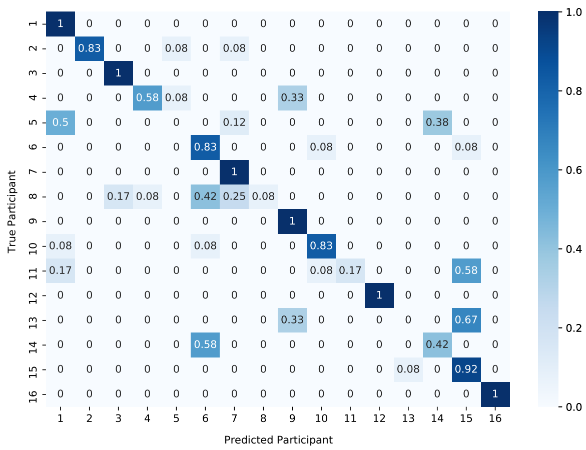
For F1 in Archery and both fused models, we again identified a significanct difference by a Friedman test ($\chi^2(3) = 11.033, p = 0.0116, N = 16$). A Wilcoxon test for *Without Normalization vs. Both Normalizations* resulted in a significant increase in identification performance ($W = 58, Z = -2.793, p < 0.025, r = 0.087$). Similarly, for Archery and F1 with two fused confusion matrices, we identified another significant Friedman test ($\chi^2(3) = 8.284, p = 0.040, N = 16$). The subsequent Wilcoxon test for *Without Normalization vs. Both Normalization* showed another significant difference ($W = 49, Z = -2.795, p < 0.025, r = 0.087$).

For feature set F2, the Friedman test only yielded a significant difference for the fused confusion matrices in Archery ($\chi^2(3) = 8.284, p = 0.040, N = 16$). The resulting Wilcoxon Test for *Without Normalization vs. Both Normalization* showed significant differences ($W = 49, Z = -2.795, p < 0.025, r = 0.087$).

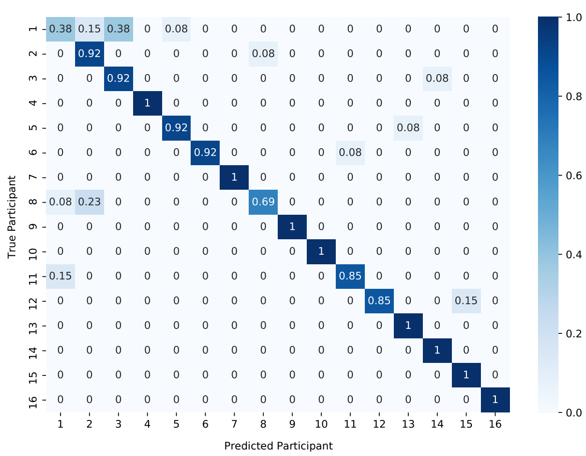
Finally, we looked into feature set F3. For Archery, when fusing the confusion matrices, a Friedman test yielded significance ($\chi^2(3) = 7.938, p = 0.047, N = 16$). None of the following Wilcoxon tests could confirm this assertion. For Bowling, the Friedman test showed, when fusing the confusion matrices, a significant result ($\chi^2(3) = 8.784, p = 0.032, N = 16$) leading to a significant Wilcoxon test in *Arm Length Normalization vs. Height Normalization* ($W = 55, Z = -2.680, p < 0.037, r = 0.084$).

An investigation of all other groups did not lead to any other statistically significant effects of body normalizations on the identification rate.

We also tested the distribution of the acquired scores (Archery: total hit pins per participant, Bowling: total hit targets per participant) for the task-based scenarios regarding the imposed body normalizations to understand whether user performance was impacted by the imposed normalization. A Friedman test showed no significant differences (Archery: ($\chi^2(3) = 0.514, p = 0.916, N = 16$), Bowling: ($\chi^2(3) = 5.905, p = 0.0116, N = 16$)).



(a) Confusion Matrix for Bowling



(b) Confusion Matrix for Archery

Figure 6: The confusion matrices for the best performing models: (a) Archery with feature set F3, imposed Height Normalization and a recurrent model and (b) Bowling for feature set F2 with Height Normalization and a recurrent model (cf., Table 1).

6.2 Interviews and TLX

After each task we asked the participants to fill out a Raw NASA TLX questionnaire (cf., Figure 7). Participants mostly rated both scenarios equally in all categories and subsequent Wilcoxon tests for each category could not show any significant differences .

In the semi-structured interview conducted after the study, participants were asked questions about their experience with the scenarios during the study as well as how they perceived the changes to their virtual body, if they noticed it at all. We asked them to rate how well they liked each scenarios on a scale from 1 to 10. The average score for Archery yielded 8.75 (SD=1.11) ; Bowling achieved with 7.93 (SD=1.22) a slightly lower score.

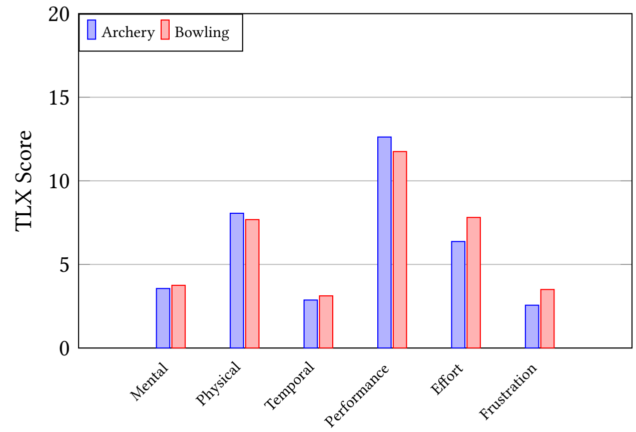


Figure 7: Results of the Raw NASA TLX that was filled out by the participant after each block in the study.

Furthermore, we asked participants if they felt that something was off or if they perceived that some factor of the environment was changing during their interaction. Out of 16 participants only four noticed some sort of change that was imposed by the given conditions. Two participants reported to have realized this change during the Bowling scenario and another two during Archery. However, three more participants mentioned that they felt the changes after we told them that we modified their body representation. In sum, less than half of the participants noticed any change.

Only one participant stated they would not like to have that kind of arm and height changes, the others did not oppose the idea. Finally, some interesting remarks were made during the study: One participant suggested they would be able to see the changes clearly if they could see the full virtual body including legs, and arms in the scenarios, not just the hands. Another participant stated that due to a lack of a reference point, it was difficult to notice a change: "I did not realize any change as I focused on the target". On the other hand, one participant noticed the change and mentioned: "I felt that I do need to bow further down to pick up the [bowling] ball".

7 DISCUSSION

We first discuss identification accuracy, continue with the impact on body normalization, and implicate the effects on deep learning.

7.1 Identification Accuracy

From the given results (cf., Table 2), it is apparent that body normalization has a strong effect on the identification rate and that identification is in general possible (RQ1). Archery peaks at a rate of 0.90 in feature set F3, followed by Bowling at a rate of 0.68 in F2. Both utilize a recurrent model. Without the imposed normalization and the same feature sets and models, the accuracies are 0.63 for Archery and 0.58 for Bowling; hence, we see an increase of 27 and 10 percentage points, respectively. This identification rate is achieved by the deep learning model validating the data of the second day of the study, while it had been trained with the data from the first day.

For Archery, only one identity was misclassified in more samples than it was classified correctly (cf., Figure 6). Given the base chance of $1/16$, i.e. ≈ 0.06 of correct classifications we overall argue, that identification across several days is reliably possible when *Height Normalization* is employed. A statistical analysis (cf., Table 3) shows several significant increases in the identification rate due to the body normalization.

We investigate the reasons for false classifications by looking into the data and video recordings of Archery. P1 and P8 were particularly hard to classify (cf., Figure 6). To gain further insight, we visualize the normalized mean heights for this task (cf., Figure 5). After also investigating the in-app recording that was created during the study, we could not find any apparent reason for the misclassification of P1. For P8, however, we found the reason for the deviations in the normalized height (compare negative outliers in Figure 5): the participant went into a crouched stance in the middle of the Archery task for three repetitions and thus shot arrows from a lower position.

7.2 Impact of Body Normalization

We believe that the normalization of the user's height forces the deep learning model to adapt its learning process. Instead of recognizing different users by their physiology (i.e., height and arm length), we force it to focus on the subtle changes in behavior between participants. Hence, we would respond to our RQ2 by estimating that physiological factors play a role in Behavioral Biometrics, but removing them leads to a reduction in noise and, therefore, to an increase in identifiability.

Although we employed several precautions, such as a subtle fading between the repetitions of the scenarios, we cannot rule out the possibility that some participants noticed this change due to its appearance at some point in the study. From the interviews in the study, we have seen that less than half of the participants noticed the body normalization. We assume that an application with a higher degree of immersion and fidelity might hide the effect even better.

7.3 Effects on Identification Systems

With regard to the feature sets that can be used for identification through deep learning models, we assume that less data yields a bigger effect. Our best result for Archery was met in feature set F3, which corresponds to only the positional coordinates and rotation of the user's controllers. As those are characterized as the vector from the controller towards the HMD, they implicitly bear information about the HMD; however, in comparison to the same result of Archery in F1 with Height Normalization in the recurrent model, the identification accuracy is increased by 21 percentage points. Here, we estimate that an overall reduction of the data in a precision task such as Archery is beneficial for the recognition rate and can result in a large positive effect. Although this does not directly translate to Bowling, the difference of only four percentage points appears minimal (i.e., Bowling in F1 with Height Normalization and the recurrent model at 0.62 vs. Bowling in F3 with Height Normalization and the recurrent model at 0.66).

8 CONCLUSION

In this work, we show that implicit identification of users in VR through their spatial motion data (which can be captured through a consumer-grade head-mounted display) is possible, given that we achieved an identification rate of up to 90% across several days. We show and evaluate identification performance in two different scenarios in a user study with 16 participants. Furthermore, we provide insight into the performance of four different feature sets and show that a reduction and normalization of data leads to a higher identification accuracy by a deep learning classifier. Moreover, we explore the concept of body normalization by virtually altering the heights and arm lengths of the users. By normalizing the body proportions of all participants, we were able to show that this improves the accuracy of the classifier. We believe that body normalization can improve future identification systems in VR.

ACKNOWLEDGMENTS

The presented work was funded by the German Research Foundation (DFG) under project no. 426052422 and 425869382 and the DAAD within the context of the ComIC project.

REFERENCES

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ Software available from tensorflow.org.

[2] Devon Adams, Alseny Bah, Catherine Barwulor, Nureli Musaby, Kadeem Pitkin, and Elissa M. Redmiles. 2018. Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX Association, Baltimore, MD, 427–442. https://www.usenix.org/conference/soups2018/presentation/adams

[3] A. Ajit, N. Banerjee, and S. Banerjee. 2019. Combining Pairwise Feature Matches from Device Trajectories for Biometric Authentication in Virtual Reality Environments. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE Computer Society, Los Alamitos, CA, USA, 9–97. https://doi.org/10.1109/AIVR46125.2019.00012

[4] F. A. Alsulaiman and A. El Saddik. 2008. Three-Dimensional Password for More Secure Authentication. *IEEE Transactions on Instrumentation and Measurement* 57, 9 (2008), 1929–1938.

[5] F. Argelaguet, L. Hoyet, M. Trico, and A. Lecuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*. IEEE, Greenville, SC, USA, 3–10. https://doi.org/10.1109/VR.2016.7504682

[6] D. A. Bowman and R. P. McMahan. 2007. Virtual Reality: How Much Immersion Is Enough? *Computer* 40, 7 (2007), 36–43.

[7] François Chollet et al. 2015. Keras. https://keras.io.

[8] I. Deutschmann, P. Nordström, and L. Nilsson. 2013. Continuous Authentication Using Behavioral Biometrics. *IT Professional* 15, 4 (2013), 12–15.

[9] Ceenu George, Daniel Buschek, Andrea Ngao, and Mohamed Khamis. 2020. GazeRoomLock: Using Gaze and Head-Pose to Improve the Usability and Observation Resistance of 3D Passwords in Virtual Reality. In *Augmented Reality, Virtual Reality, and Computer Graphics*, Lucio Tommaso De Paolis and Patrick Bourdot (Eds.). Springer International Publishing, Cham, 61–81.

[10] Ceenu George, Mohamed Khamis, Emanuel von Zezschwitz, Henri Schmidt, Marinus Burger, Florian Alt, and Heinrich Hu§mann. 2017. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *Proceedings 2017 Workshop on Usable Security*. Internet Society, San Diego, CA, USA. https://doi.org/10.14722/usec.2017.23028

[11] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[12] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in*

Psychology. Elsevier, Amsterdam, The Netherlands, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9

[13] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation* 9, 8 (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[14] K. Igarashi, C. Miyajima, K. Itou, K. Takeda, F. Itakura, and H. Abut. 2004. Biometric identification using driving behavioral signals. In *IEEE International Conference on Multimedia and Expo (ICME)*, Vol. 1. IEEE, Taipei, Taiwan, 65–68 Vol.1. https://doi.org/10.1109/ICME.2004.1394126

[15] Markus Jakobsson, Elaine Shi, Philippe Golle, and Richard Chow. 2009. Implicit Authentication for Mobile Devices. In *Proceedings of the 4th USENIX Conference on Hot Topics in Security* (Montreal, Canada) *(HotSec'09)*. USENIX Association, USA, 9.

[16] Hassan Khan, Urs Hengartner, and Daniel Vogel. 2016. Targeted Mimicry Attacks on Touch Input Based Implicit Authentication Schemes. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (Singapore, Singapore) *(MobiSys '16)*. Association for Computing Machinery, New York, NY, USA, 387–398. https://doi.org/10.1145/2906388.2906404

[17] Hassan Khan, Urs Hengartner, and Daniel Vogel. 2018. Augmented Reality-Based Mimicry Attacks on Behaviour-Based Smartphone Authentication. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (Munich, Germany) *(MobiSys '18)*. Association for Computing Machinery, New York, NY, USA, 41–53. https://doi.org/10.1145/3210240.3210317

[18] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (2012), 373–387. https://doi.org/10.1162/PRES_a_00124

[19] Konstantina Kilteni, Jean-Marie Normand, Maria V. Sanchez-Vives, and Mel Slater. 2012. Extending Body Space in Immersive Virtual Reality: A Very Long Arm Illusion. *PLOS ONE* 7, 7 (07 2012), 1–15. https://doi.org/10.1371/journal.pone.0040867

[20] Alexander Kupin, Benjamin Moeller, Yijun Jiang, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Task-Driven Biometric Authentication of Users in Virtual Reality (VR) Environments. In *MultiMedia Modeling*, Ioannis Kompatsiaris, Benoit Huet, Vasileios Mezaris, Cathal Gurrin, Wen-Huang Cheng, and Stefanos Vrochidis (Eds.). Springer International Publishing, Cham, 55–67.

[21] Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. 2016. Data Augmentation for Time Series Classification using Convolutional Neural Networks. In *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*. Riva Del Garda, Italy. https://halshs.archives-ouvertes.fr/halshs-01357973

[22] Marta Matamala-Gomez, Tony Donegan, Sara Bottiroli, Giorgio Sandrini, Maria V. Sanchez-Vives, and Cristina Tassorelli. 2019. Immersive Virtual Reality and Virtual Embodiment for Pain Relief. *Frontiers in Human Neuroscience* 13 (2019), 279. https://doi.org/10.3389/fnhum.2019.00279

[23] Florian Mathis, Hassan Ismail Fawaz, and Mohamed Khamis. 2020. Knowledge-Driven Biometric Authentication in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3334480.3382799

[24] Cameron Appel Max Roser and Hannah Ritchie. 2013. Human Height. https://ourworldindata.org/human-height, retrieved 2021-01-11.

[25] Lukas Mecke, Daniel Buschek, Mathias Kiermeier, Sarah Prange, and Florian Alt. 2019. Exploring Intentional Behaviour Modifications for Password Typing on Mobile Touchscreen Devices. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA. https://www.usenix.org/conference/soups2019/presentation/mecke-behaviour

[26] R. Miller, A. Ajit, N. Kholgade Banerjee, and S. Banerjee. 2019. Realtime Behavior-Based Continual Authentication of Users in Virtual Reality Environments. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. IEEE, San Diego, CA, USA, 253–2531. https://doi.org/10.1109/AIVR46125.2019.00058

[27] R. Miller, N. K. Banerjee, and S. Banerjee. 2020. Within-System and Cross-System Behavior-Based Biometric Authentication in Virtual Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Atlanta, Georgia, USA, 311–316. https://doi.org/10.1109/VRW50115.2020.00070

[28] Tahrima Mustafa, Richard Matovu, Abdul Serwadda, and Nicholas Muirhead. 2018. 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* (Tempe, AZ, USA) *(IWSPA '18)*. Association for Computing Machinery, New York, NY, USA, 23–30. https://doi.org/10.1145/3180445.3180450

[29] Ilesanmi Olade, Charles Fleming, and Hai-Ning Liang. 2020. BioMove: Biometric User Identification from Human Kinesiological Movements for Virtual Reality Systems. *Sensors* 20, 10 (2020), 2944.

[30] Panagiota Papadopoulou. 2007. Applying virtual reality for trust-building e-commerce environments. *Virtual Reality* 11, 2-3 (2007), 107–127.

[31] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300340

[32] Henning Pohl, Markus Krause, and Michael Rohs. 2015. One-Button Recognizer: Exploiting Button Pressing Behavior for User Differentiation. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 403–407. https://doi.org/10.1145/2750858.2804270

[33] Sarah Prange, Daniel Buschek, Ken Pfeuffer, Lukas Mecke, Peter Ehrich, Jens Le, and Florian Alt. 2019. Go for GOLD: Investigating User Behaviour in Goal-Oriented Tasks. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312949

[34] Albert "Skip" Rizzo and Gerard Jounghyun Kim. 2005. A SWOT analysis of the field of virtual reality rehabilitation and therapy. *Presence: Teleoperators & Virtual Environments* 14, 2 (2005), 119–146.

[35] Sylvia Rothe, Boris Kegeles, and Heinrich Hussmann. 2019. Camera Heights in Cinematic Virtual Reality: How Viewers Perceive Mismatches Between Camera and Eye Height. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Salford (Manchester), United Kingdom) *(TVX '19)*. Association for Computing Machinery, New York, NY, USA, 25–34. https://doi.org/10.1145/3317697.3323362

[36] Albrecht Schmidt. 2000. Implicit human computer interaction through context. *Personal technologies* 4, 2 (2000), 191–199.

[37] Neal E Seymour, Anthony G Gallagher, Sanziana A Roman, Michael K O'brien, Vipin K Bansal, Dana K Andersen, and Richard M Satava. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery* 236, 4 (2002), 458.

[38] M. Sivasamy, V. N. Sastry, and N. P. Gopalan. 2020. VRCAuth: Continuous Authentication of Users in Virtual Reality Environment Using Head-Movement. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, Coimbatore, India, 518–523. https://doi.org/10.1109/ICCES48766.2020.9137914

[39] Chee Meng Tey, Payas Gupta, and Debin Gao. 2013. I can be you: Questioning the use of keystroke dynamics as biometrics. In *Annual Network and Distributed System Security Symposium 20th NDSS 2013, February*. Internet Society, San Diego, CA United States, 24–27.

[40] Z. Yu, H. Liang, C. Fleming, and K. L. Man. 2016. An exploration of usable authentication mechanisms for virtual reality systems. In *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*. IEEE, Jeju, South Korea, 458–460. https://doi.org/10.1109/APCCAS.2016.7804002