A Systematic Investigation of Device Combinations and Spatial Representations for Identifying Virtual Reality Users

Alec G. Moore* University of Central Florida
Tiffany D. Do† University of Central Florida
Nicholas Ruozzi‡ University of Texas at Dallas
Ryan P. McMahan§ University of Central Florida

Abstract

Recently, there has been much interest in using virtual reality (VR) tracking data to authenticate or identify users. In this paper, we present one of the first investigations of how different combinations of VR input devices (e.g., headset, dominant hand controller, offhand controller) and their spatial representations (e.g., position and/or rotation via Euler angles, quaternions, or 6D) affect identification accuracy. To facilitate this investigation, we conducted a user study ($n = 45$) involving participants learning how to assemble two distinct full-scale constructions. Our results indicate that the availability of more tracked devices improve identification accuracies for the same assembly task, but only the headset affords the best accuracies for a similar task. Our results also indicate that spatial features involving position and any rotation representation yield better accuracies than either alone. Finally, we demonstrate that first-order derivatives can be used to obfuscate user identities for privacy concerns.

Index Terms: Human-centered computing—Virtual reality; Security and privacy—Privacy protections

1 Introduction

There have been several works recently that investigate the identifiability of virtual reality (VR) users based on properties of their biomarkers. Some works, such as Pfeuffer et al. [15], focus on an authentication model, where the user proposes their identity, and the system verifies that they are indeed that person. Other works, like those by Miller et al. [7] and Moore et al. [10] look at the question of passive identifiability. That is, given a person's biomarkers data in a population, can they be identified afterwards from a new sample of their VR usage data.

While the question of identifiability has been explored in several contexts, recent work by Liebers et al. [5], examines identifiability within a gameified context. Other environments, such as the one used by Asish et al. [1], are built as educational experiences, while some are developed specifically to elicit identifiable motions like one used in a paper by Liebers et al. [6]. In work by Mustafa et al. [12], the authors recognize that their approaches may yield different results when applied to a different context.

We expand upon a shortcoming of prior works by providing an analysis into the identifiability of users across two separate VR training sessions, where both samples are captured within a single hour span of time. Prior work by Miller et al. [7] found very high identification rates, they were between single-session samples of data. The work by Moore et al. [9], on the other hand, showed lower identification rates, but between sessions collected a week apart, introducing potential confounds due to the participants' mental state, clothes, physical health, and other potential changes. By collecting data within the same period of time, but separate VR sessions, we can isolate and begin to understand to what degree identifiability diminishes as a result of exiting and reentering VR.

In this work, we choose a simple training task as our ecologically valid environment since this domain has seen expanded use as of late [18]. We developed a virtual environment to train users how to construct simple objects out of a set of toy pipes and connectors, allowing us to emulate a simple assembly task. By using a prescribed set of instructions that the user had to replicate, we reduce the likelihood that our within-session models overfit to features of the participant's experience, because we know that all participants completed the same steps in the same order for the same tasks.

Using this training task, we conducted a study with 45 participants, yielding 1.7 million frames of data over more than 5 hours. As far as research on identifiability with VR experiences this amounts to less than the 60 participants of data Moore et al. [9] had, and far fewer than the 511 participants of the work by Miller et al. [7]. Unlike those works, however, our informed consent allows us to publish this dataset to make it openly available for future research.

Finally, we conduct 4 machine learning experiments to examine the identifiability of this data, moderating the inclusion of data, its representation and the models, yielding results for 1176 total conditions. We choose to examine only the machine learning models Random Forest (RF), Gradient Boosting Machine (GBM), and k-nearest neighbors (kNN) rather than deep learning models due to their reduced likelihood of overfitting. In our best-performing within-session RF model, we find over 95% accuracy, but that same condition drops in accuracy to around 46% when trained on one session and evaluated on another. We also examine a set of featurizations based on the 1st order time derivative of the data and find it to perform moderately well, but with a different set of data inclusion.

In this paper, we will discuss some works related to our efforts, describe virtual environment and the experimental methodologies, discuss our findings, and finally conclude with the limitations of this work and some future directions. The primary contributions presented herein are:

1. An analysis of identifiability between VR sessions within a short span of time.
2. An exhaustive exploration of data inclusion and representation conditions and how they affect identifiability.
3. We provide a novel dataset featuring high-framerate capture of the tracked devices across two VR sessions for 45 participants. A dataset containing all data used in this paper will be made available at GitHub: [LINK REDACTED FOR REVIEW]

2 Related Works

2.1 Identification and Authentication with Eye-tracking

Several commercially available HMDs for augmented or virtual reality affort eye-tracking. Often the hardware necessary for this is included as a means of collecting user input or improving the hardware performance through techniques like foveated rendering. As this stream of data has become more available, several researchers have begun exploring its usefulness as a means to identify or authenticate users.

*e-mail: agm@knights.ucf.edu
†e-mail: tiffanydo@knights.ucf.edu
‡e-mail: nicholas.ruozzi@utdallas.edu
§e-mail: rpm@ucf.edu

Table 1: A comparisons of the input features used by related works. Each letter under a given paper indicates a condition that was evaluated in that paper consisting of the data for each row in which the letter appears. The chart also shows what kinds of classifiers each work explored as well as the number of participants, and the highest accuracy attained by their best-performing classifier among the explored representations.

			[2]	**[6]**	**[17]**	**[13]**	**[1]**	**[12]**	**[7]**	**[16]**	**[5]**	**[9]**	**[15]**	**Ours (Pos)**	**Ours (Vel)**
Input Features	**Eye**	**2D Pos**	A	A		A									
		3D Dir					A								
		Others	A		A		A								
	Head	**Position**			AB	A			A	A	A	A	AEP	AGJK PQSU	
		Velocity										B	BEP		AGJK PQSU
		Distance											CEP		
		Rotation		A	AB	A		A	A	AB	A	A	DEP	BGMN PQTU	
		Ang Vel								C		B			BGMN PQTU
	Dominant Hand	**Position**			A	A			A	AB	AB	A	FJP	CHJL PRSU	
		Velocity								C		B	GJP		CHJL PRSU
		Distance											HJP		
		Rotation			A	A			A	AB	AB	A	IJP	DHMO PRTU	
		Ang Vel								C		B			DHMO PRTU
	Off Hand	**Position**			A	A			A	AB	AB	A	KOP	EIKL QRSU	
		Velocity								C		B	LOP		EIKL QRSU
		Distance											MOP		
		Rotation			A	A			A	AB	AB	A	NOP	FINO QRTU	
		Ang Vel								C		B			FINO QRTU
Classifiers			RBF	DNN kNN	LogR DT, RF	kNN	CNN kNN LSTM RF	LogR SVM	GBM kNN RF	RF, MLP FRNN LSTM GRU	MLP RNN	GBM kNN RF	RF SVM	GBM kNN RF	GBM kNN RF
Participants			18	12	35	15	34	23	511	34	16	60	22	45	45
Highest Accuracy			85%	100%	96%	98%	98%	93%	95%	100%	90%	90%	44%	96%	68%

In recent work by David-John et al. [2], the authors recognize that while eye-tracking data can be useful as a means of input, it also runs a heightened risk of allowing systems identify users, for example, by correlating a users' data across multiple accounts. They propose a privacy-preserving means of streaming eye-tracking data by gatekeeping at the API level. Their approach was capable of reducing identifiability from 85% to approximately 30% while preserving a system's ability to use gaze data for input like foveated rendering.

In another recent work by Liebers et al [6], the authors make use of eye-tracking features and HMD orientation and provided participants with a stimulus designed to elicit smooth pursuit head and eye movements. These movements were tracked by logging the reported HMD Euler angles, as well as the pupil position. After preprocessing to determine additional eye-tracking events such as saccades and pursuits, The authors then investigate both kNN as well as a set of 10 Deep Learning Neural Network approaches. Ultimately the authors found that inclusion of HMD-based data increased accuracy of their ML model from 45% to 90%, as well as their best-performing deep learning model from 96% to 100%.

Expanding on the number of tasks examined, Tricomi et al. [17] present a pre-print investigating the identifiability of participants in both VR and AR tasks. The participants were exposed to 5 types of tasks in the AR condition and 7 types of tasks in the VR condition. They make use of automated systems to determine salient features of the raw data, then use machine learning to attempt to identify and profile their participants. The authors found 96% identifiability accuracy.

Olade et al. [13] also examined similar body and eye tracking features, but for both continuous identification as well as authentication. They used a data set of 15 participants, and investigated various attack types could one could be conducted and what the risk was among their participants. Ultimately, their accuracy was 98.6%.

Finally, one more paper that investigates identifiability with the inclusion of eye-tracking data by Asish et al., [1] focused exclusively on eye-tracking data. This paper has VR session divided into 4 separate experiences and examines identifiability across those sessions. The authors also train their models on 3 of the 4 experienced sessions and find 98% overall accuracy with their deep learning models.

2.2 Identification and Authentication without Eye-tracking

While eye-tracking has been incorporated in multiple HMDs, many headsets do not support it as a form of input. Several works have investigated identifiability or authentication by making use only of the sensors that provide the tracking needed for a virtual reality experience. The benefit to restricting oneself to this set of data is that the results are applicable to a broader range of hardware than only those that afford eye-tracking.

Mustafa et al. [12] explore authentication in a VR experience making use of only features derived from the orientation of the HMD. In their work, they generated authentication models that gave equal-error rates around 7%, showing feasibility in such an approach for authenticating in a scenario with 23 users. One caveat pointed out by the authors is that their results were task-specific due to encoding features of the virtual environment, and so authentication in another environment would require the creation of new models which may perform differently.

In recent work by Miller et al. [7], the authors presented a study involving 511 participants with a virtual environment that displayed 360° video clips with questionnaires presented between. They made use of only the positions and orientations of the controllers and head-mounted display for passive identification of users. They conducted an ablation study examining the removal of subcomponents of tracking data and found that removal of the HMD Y value (which is strongly correlated with height) resulted in the largest drop in identification accuracy.

Schell et al. [16] also investigate identifiability on motion from a restricted set of data points. In this pre-print, the authors investigated identifiability by using the open Talking with Hands dataset [4]. Schell et al. examined creating head-relative values for the hand position and orientations, as well as their time-derivatives from this conversational dataset. This subset of the data was chosen as it corresponds to the tracking data that a typical room-scale system affords. They found that with majority voting on increasingly long test sequences, they were ultimately able to attain 100% accuracy with several of their explored models. While the data this analysis is conducted on is from in-person human-human conversational dyads outside of VR, the authors suggest that their results further contribute to the body of research showing potential in the use of this data for identifying individuals in VR experiences.

In another recent work by Liebers et al. [5], the authors investigate the identifiability of users performing two different tasks in VR. In their study, they had 16 participants perform repetitions of prescribed tasks over two days. By training their models on data from a single day's session and evaluating on a different day, they ensured that their models weren't encoding data that may be session-specific. They found that for their, they were able to attain accuracy up to 90% with a motion mapped to a normalized human body model with a Recursive Neural Network based on LSTMs and a Multilayer Perceptron.

Another work that investigated identifiability by Moore et al. [9] examined the passive identification of users across a week delay between sessions in an interactive training environment. They found that the identification accuracy across sessions was greatly reduced from around 90% to near 32%. The authors hypothesize a few reasons for this finding including variability in the presentation of the VR experience, the potential for the user to be in a substantially different physical and mental state, and wearing different clothing. The authors additionally examined using velocity-based features and found that those further reduced identification accuracy.

Finally, Pfeuffer et al. [15] also investigate identification between sessions with a minimum 3-day period between exposure. The examined identification making use of features derived from the head, hand, eye, among 22 participants. They also looked at 4 different types of simple interactions to identify users with. With multiple samples of each atomic interaction, they were able to achieve 44% identification accuracy between sessions.

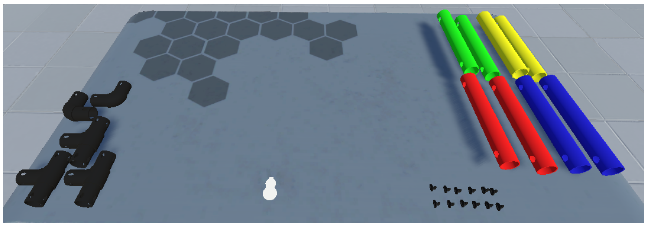

Figure 1: Both builds made use of the same pieces, in the same locations. They consisted of two of each color pipe (red, green, blue, yellow), two elbow connectors, three three-way connectors, and 12 screws. Also visible in the middle is the metallic key used to turn the screws to secure connections.

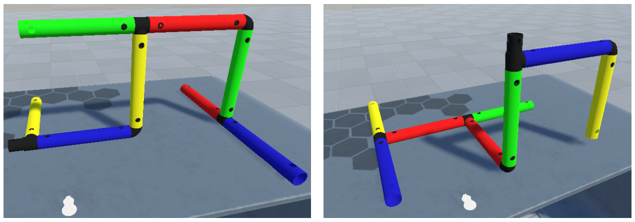

Figure 2: The completed structures. Build A is shown on the left, and build B on the right. Screws do not appear as an artifact of the way the screenshots were captured, but were visible in the application.

3 User Study

For this study, we designed a system in Unity to train a user how to build an arbitrary object using a toy set of pipes and connectors. We identified two structures to teach and created applications that corresponded to those structures. Throughout this paper, we will refer to these structures/applications as "A" or "B". Each arbitrary object made use of the same set of pieces, as shown in Figure 1. The steps for the assembly order was prescribed according to Table 2, with users beginning with a connector, attaching a pipe, then using the key to screw in a screw to secure the connection.

In order to know which piece to attach, where, and with what orientation, we employed an animated interaction cue to guide the user to complete the current step [3]. This interaction cue was presented as a transparent copy of either the controller representation or the held object, and smoothly animated from its current location to where it needed to be. If an object was currently being held and needed to be attached, the system used the copy of the object. Similarly, if an object needed to be grabbed, it would show an animation from the closest controller that wasn't currently holding any object.

Additional pipes and connectors were then added to the existing structure with participants having to use a screw to secure each

Build	1	2	3	4	5	6	7	8	9	10	11	12
A	Y1	B1	L1	Y2	T2	G1	R1	L2	G2	T3	R2	B2
	T1	T1	B1	L1	Y2	T2	T2	R1	L2	G2	T3	T3
B	Y1	B1	R1	T2	G1	R2	L1	G2	T3	B2	L2	Y2
	T1	T1	T1	R1	T2	T2	R2	L1	G2	T3	B2	L2

Table 2: The order of steps for builds A and B. The letters R, G, B, and Y represent red, green, blue, and yellow pipes respectively, T and L represent Three-way and Elbow joints, respectively.

Table 3: An overview of the conditions compared in this paper. Explored conditions include the 0th and 1st order time derivative, within- and between-session predictions, the inclusion and exclusion of individual trackers, the type of data used from those trackers, and the model used.

Trackers Used				Included Data			Model Type		Session		Time Derivative
Head	DomH	OffH		Position	Euler						
Head	DomH			Position	Quaternion				A → A		
Head		OffH		Position	SixD		RF		B → B		0th
	DomH	OffH	×	Position		×	GBM	×	A → B	×	1st
Head					Euler		kNN		B → A		
	DomH				Quaternion						
		OffH			SixD						

connection progressively. Upon starting the application, the pieces were frozen in place with disabled physics and unable to be grabbed until they became relevant for the current assembly step. Whenever a piece or the key fell off the table, its position was reset to the starting position after a brief period. Likewise, if the structure fell off of the table, its position would be reset to the middle of the table.

In addition to the two primary A and B assembly applications, a third with additional scaffolding was built in order to train participants how to interact with the VR system. This version of the application featured audio and text detailing which buttons to press on the controller, how to grasp an object, and had participants assemble a simple model consisting of fewer pieces than the two sessions used for this evaluation.

3.1 Materials

We made use of the HTC Vive Pro Eye VR system to run this study. Retraction cables connected to the ceiling were attached to the cable of the head-mounted display to ensure free movement of the participants and reduce the chance of a trip hazard. To run the benchmark application, we made use of a PC with an NVidia GeForce RTX 2080 graphics card. The application ran at a consistent 90fps.

3.2 Procedure

The following experiment was approved by the University Institutional Review Board. A link to a pre-survey was made available through university mailing lists. The pre-survey first ensured that people responding to our survey were eligible according to our inclusion criteria. If so, they then were then asked for their consent to participant, and finally the pre-survey collected their demographics. Participants were then invited to schedule a 1 hour period of time to complete the in-person portion of the experiment. Upon the participant's arrival, the experimenter collected informed consent. Participants were then exposed to the tutorial application, followed by a break with a questionnaire. After completing the first questionnaire break, participants were either exposed to the A or B assembly task, depending on their cohort to ensure counterbalancing of order of presentation. Participants then had an additional break with questionnaires, followed by the assembly task they had not yet done. Participants concluded with a final set of exit questionnaires.

For each questionnaire break, the experimenter would assist the participant in removing the head-mounted display, collect their controllers, and administer questionnaires presented on a 2-dimensional monitor using a mouse and keyboard as input.

3.3 Participants

A total of 45 participants were recruited via university mailing lists. All participants (20 females, 25 males) had normal or corrected-to-normal vision with contacts, which were worn through the duration of the study. The mean age of our participants was 22.1 ± 4.2, and 3 were left-hand dominant.

Position A → A RF	**Position + Euler**	**Position + Quaternion**	**Position + SixD**	**Position**	**Euler**	**Quaternion**	**SixD**
Head + DomH + OffH	95.1%	95.3%	95.8%	85.6%	85.9%	85.4%	87.8%
Head + DomH	94.7%	93.4%	94.3%	78.3%	74.4%	73.0%	75.6%
Head + OffH	93.9%	93.7%	94.3%	79.7%	76.6%	76.8%	77.6%
DomH + OffH	88.7%	87.9%	88.4%	64.4%	71.0%	72.8%	71.3%
Head	85.1%	85.0%	86.0%	53.2%	42.9%	40.7%	44.2%
DomH	66.6%	67.0%	67.4%	41.7%	46.7%	45.7%	45.7%
OffH	71.2%	72.2%	71.9%	39.4%	49.9%	45.7%	47.8%

Table 4: Within-session identification accuracy for Random Forest, with position and orientation data, trained and evaluated with data from session A.

4 Machine Learning Experiment 1

In this machine learning experiment, we analyze the identifiability of motion data within each session. That is to say that we make use of data from a given session for training our models, and some retained data for evaluating their accuracy. Several works explore the identifiability of users within a single VR session, such as that by Miller et al. [7].

For each session the participants experienced, the system tracked the position and orientation of the HMD and controllers at a rate of 90Hz. Because the application was built in the Unity Engine for the HTC Vive Pro Eye, this was the framerate at which the application executed its event loop, and thus the rate at which position and orientation data was provided to Unity to ensure the application updated the rendered view for the HMD.

Across all analyzed conditions, we considered the inclusion or exclusion of each tracked object (the HMD on the head, the controller in the dominant hand, and the controller in the non-dominant hand), as well as inclusion of position or orientation. Although the inclusion and exclusion of each individual value from each tracked object could be toggled independently (such as including or excluding the Y component of position), the combination of these features would result in an untenable search space. We choose to moderate the inclusion and exclusion of data by tracked object to allow us to explore how each tracked object is contributing to identifiability. Likewise, we moderate the inclusion of position and orientation data so we can evaluate if either are too noisy for the models to fit well.

Beyond the inclusion of orientation data, we also considered three orientation representations: Euler angles, quaternions, and a six-dimensional representation. While the authors are not aware

Position B → B RF	Position + Euler	Position + Quaternion	Position + SixD	Position	Euler	Quaternion	SixD
Head + DomH + OffH	94.2%	95.7%	95.0%	81.8%	83.1%	83.0%	84.1%
Head + DomH	91.8%	91.0%	91.0%	72.2%	70.6%	68.6%	69.3%
Head + OffH	94.4%	93.2%	94.4%	77.3%	75.7%	75.3%	76.6%
DomH + OffH	86.0%	87.3%	85.9%	59.6%	71.8%	73.8%	70.8%
Head	84.7%	81.3%	83.8%	48.0%	37.4%	36.9%	40.1%
DomH	61.6%	61.2%	59.7%	32.2%	47.3%	49.2%	49.4%
OffH	71.0%	71.7%	70.4%	40.1%	56.0%	56.2%	51.8%

Table 5: Within-session identification accuracy for Random Forest, with position and orientation data, trained and evaluated with data from session B.

of a six-dimensional representation being applied specifically to data collected in VR experiences, this rotation representation has recently shown promise for machine learning models because of the lack of discontinuities among continuous data [19]. We also examined Euler angles and quaternions as alternatives due to their ready availability in the Unity engine and to help with comparisons with prior works [8] [12].

The models we chose to evaluate were k-Nearest Neighbors, Random Forests, and Gradient-Boosting Machine. These have been used previously by Miller et al. [7] effectively for identifying participants. Because of the large number of hyperparameters we are already evaluating across we choose to make use of default intrinsic parameter values for each model in the Scikit-Learn library [14]. These values were k=3 for kNN, 100 estimators for Random Forest, and 100 estimators and a learning rate of 0.1 for GBM. For kNN, training data was normalized and the same scaling function was applied to the evaluation data to avoid issues with the scales along different axes. We chose not to examine deep learning models due to the likelihood of overfitting to the relatively short samples of data from each participant.

The full set of hyperparameter conditions explored in this paper are described in Table 3. For this section, we discuss the results of moderating the trackers used, included data, model type and the first two values in the session column.

For a given set of conditions, we computed a set of per-second feature vectors. This feature vector described per-second statistics of each available field in the data by calculating the minimum, maximum, mean, median, and standard deviation. As an example, if head position was included in the raw data, a corresponding feature vector would include these statistics for the x, y, and z values. These feature vectors from each session were then partitioned into 10 subsessions. For the within-session identifiability analyses, we evaluated 20 Monte-Carlo shuffles of the data by training on 9 random subsessions and evaluating the accuracy on the remaining subsession. While the subsessions retained for evaluation for each participant varied within shuffles, the shuffles themselves remained consistent across conditions. The models would predict a participant ID label per 1-second feature vector provided. These 1-second level predictions were aggregated together to yield a final predicted label for a given participant's evaluation data by selecting the most-predicted label.

Position A → B RF	Position + Euler	Position + Quaternion	Position + SixD	Position	Euler	Quaternion	SixD
Head + DomH + OffH	51.1%	44.4%	46.7%	42.2%	46.7%	40.0%	44.4%
Head + DomH	53.3%	53.3%	48.9%	40.0%	35.6%	40.0%	33.3%
Head + OffH	57.8%	46.7%	51.1%	44.4%	37.8%	40.0%	40.0%
DomH + OffH	33.3%	33.3%	35.6%	22.2%	33.3%	33.3%	31.1%
Head	75.6%	71.1%	75.6%	42.2%	40.0%	31.1%	31.1%
DomH	31.1%	35.6%	31.1%	20.0%	31.1%	28.9%	26.7%
OffH	26.7%	24.4%	31.1%	15.6%	17.8%	17.8%	22.2%

Table 6: Between-session identification accuracy for Random Forest, with position and orientation data, trained on A and evaluated on B.

Position B → A RF	Position + Euler	Position + Quaternion	Position + SixD	Position	Euler	Quaternion	SixD
Head + DomH + OffH	51.1%	51.1%	53.3%	40.0%	46.7%	46.7%	44.4%
Head + DomH	57.8%	51.1%	55.6%	40.0%	42.2%	40.0%	40.0%
Head + OffH	62.2%	62.2%	62.2%	37.8%	40.0%	40.0%	40.0%
DomH + OffH	40.0%	37.8%	37.8%	31.1%	33.3%	31.1%	33.3%
Head	80.0%	75.6%	82.2%	33.3%	51.1%	31.1%	42.2%
DomH	35.6%	35.6%	35.6%	15.6%	31.1%	31.1%	24.4%
OffH	31.1%	31.1%	31.1%	22.2%	24.4%	22.2%	22.2%

Table 7: Between-session identification accuracy for Random Forest, with position and orientation data, trained on B and evaluated on A.

For positional data, within-sessions, we found our best performance among conditions including both the head and at least one hand, as well as incorporating both positional and rotational data (Figure 4,5). For session A, this was 95.8% accuracy with the RF model, all three trackers, using both position and 6-dimensional rotations. The same condition performed the best with the GBM model at 95.7% accuracy, and the kNN model at 90.1%. For session B, the best performance was also with all three trackers, but with position and quaternion rotations, at 95.7%. Again this condition was the best for GBM as well at 94.0% accuracy, for kNN the best was all three trackers with position and 6-dimensional rotations at 87.4% accuracy.

5 Machine Learning Experiment 2

For our next machine learning experiment, we investigate the between-session accuracy of these models. This is motivated by some prior works that indicate that the task of identifying an individual across multiple sessions is more difficult than within one session. Miller et al. [7] mention it as a limitation of their study, and while Liebers et al. [5], Moore et al. [11], and Pfeuffer et al. [15] all make use of datasets involving multiple sessions, the sessions are spaced at least a day apart, introducing some additional variability due to the change in state of the participant.

For our Between-session Identifiability analyses, we evaluate identification accuracy across the same set of variables as the previous section. We moderate the inclusion of trackers, the position and orientation data of those trackers as well as the orientation representation, and model type, as shown in Table 3. In this section, we now look at training models on one session and evaluating on another.

For this analysis, we developed the same per-second feature vec-

Velocity A → A RF	Position + Euler	Position + Quaternion	Position + SixD	Position	Euler	Quaternion	SixD
Head + DomH + OffH	54.2%	59.0%	64.1%	41.3%	45.7%	52.4%	62.0%
Head + DomH	48.6%	52.7%	60.3%	37.4%	40.4%	46.4%	54.8%
Head + OffH	44.8%	47.9%	53.8%	30.7%	36.2%	42.8%	48.0%
DomH + OffH	49.4%	56.0%	59.6%	39.3%	40.8%	51.0%	54.9%
Head	29.6%	32.2%	36.1%	18.3%	24.4%	28.1%	28.4%
DomH	42.4%	45.1%	49.1%	25.8%	32.3%	38.6%	43.7%
OffH	36.1%	45.1%	50.1%	27.7%	26.7%	37.7%	44.8%

Table 8: Within-session identification accuracy for Random Forest, with velocity and angular velocity data, trained and evaluated on A.

tors as previously described. Our models were then trained on the entirety of a given session and evaluated on the entirety of the other session, aggregating predictions at the second level to create a prediction for that session. This approach was chosen as it would be an ecologically valid approach for identifying users without their knowledge.

Looking at the between-session positional data, we now find our best performances still incorporated positional and rotation data, but now only made use of the head tracker (Table 6,7). For the models trained on A and evaluated on B, the best performance was head tracker, position, and either Euler or 6-dimensional rotational representation, at 75.6% for RF, 73.3% for GBM, and 75.6% for kNN. For the models trained on B and evaluated on A, the head tracker and position with 6-dimensional rotation performed at 82.2% for RF, and 80% for GBM. For B to A, kNN performed best with head, position and Euler angles at 71.1%.

6 Machine Learning Experiment 3

Beyond creating the feature vector by using the collected data as it was, we also considered computing its first-order time derivative, since velocity-based features have been demonstrated to be useful when predicting knowledge and performance retention [11]. This tracking data was then aggregated at the one-second level by computing the minimum, maximum, mean, median, and standard deviation for each component value, to create a feature vector to describe that second of motion. Depending on the tracker inclusion and position/orientation conditions, this per-second feature vector consisted of between 15 and 135 values.

In this section, we again divided the data from each session into 10 subsessions. We use a similar approach of creating 20 Monte-Carlo shuffles of the subsessions, allowing the models to train on 9 of them per participant, and evaluating their identification accuracy on the retained ones. This is essentially the same procedure as presented in Section 4, but now conducted on the feature vectors that were generated using the time-derivative data instead of the raw data. Again, RF and GBM performed on par with each other, and overall better than kNN, so we choose to show the results for RF (Table 8,9), with the full set of results available in the supporting material.

Examining the within-session velocity data, we found our highest identification accuracies among models involving six-dimensional rotation representations and incorporating at least two trackers. Across all models, for both the A session and B session, the best-performing conditions were those that made use of position and 6-dimensional rotations across all three trackers, yielding 64.1% accuracy for session A and 66.2% accuracy for session B. GBM maxed out at 64.3% for session A and 68.0% for B. Finally, kNN performed notably worse, with accuracies of 37.2% for A and 40.9% for B. k

Velocity B → B RF	Position + Euler	Position + Quaternion	Position + SixD	Position	Euler	Quaternion	SixD
Head + DomH + OffH	58.7%	63.6%	66.2%	46.3%	53.0%	59.7%	64.2%
Head + DomH	51.8%	54.2%	58.3%	35.1%	41.7%	48.4%	54.6%
Head + OffH	50.9%	55.7%	60.8%	36.9%	41.4%	50.9%	56.3%
DomH + OffH	52.1%	58.7%	60.9%	36.7%	44.6%	53.4%	56.9%
Head	30.2%	31.9%	36.6%	20.2%	22.3%	25.6%	31.2%
DomH	41.6%	44.2%	46.4%	24.1%	27.4%	39.1%	43.2%
OffH	40.6%	49.0%	49.9%	26.2%	29.1%	40.1%	42.7%

Table 9: Within-session identification accuracy for Random Forest, with velocity and angular velocity data, trained and evaluated on B.

Velocity A → B RF	Position + Euler	Position + Quaternion	Position + SixD	Position	Euler	Quaternion	SixD
Head + DomH + OffH	28.9%	28.9%	35.6%	17.8%	26.7%	28.9%	33.3%
Head + DomH	28.9%	24.4%	33.3%	15.6%	20.0%	28.9%	31.1%
Head + OffH	28.9%	31.1%	24.4%	15.6%	22.2%	22.2%	24.4%
DomH + OffH	22.2%	22.2%	26.7%	13.3%	17.8%	22.2%	24.4%
Head	33.3%	42.2%	44.4%	15.6%	28.9%	28.9%	33.3%
DomH	20.0%	20.0%	24.4%	15.6%	11.1%	11.1%	24.4%
OffH	24.4%	24.4%	20.0%	11.1%	20.0%	17.8%	15.6%

Table 10: Between-session identification accuracy for Random Forest, with velocity and angular velocity data, trained on data from A and evaluated with data from session B.

7 Machine Learning Experiment 4

Finally, we examine the identifiability of motion data, using the same velocity feature-vector procedure described in the previous section, but now between sessions. We conducted this exploration by allowing the models to train over the entirety of one session from all participants, then evaluated over the entirety of the other session. Again, because our models created predictions for each second, based on each per-second feature vector provided, we aggregate the predictions, saying that the model ultimately predicted an identity based on

Finally, when we look at the between-session velocity data, we find that no conditions exceeded 50% accuracy. Our best-performing models involved both the position and 6-dimensional representation of the data, with the head tracker only. For the RF model trained on session A and evaluated on session B, we found 44.4% accuracy, and for the RF model trained on B and evaluated on A, this yielded 46.7% accuracy. For that same condition, GBM had an accuracy of 46.7% for A to B. For GBM's best B to A condition, we see head and dominant-hand, position and 6-dimensional rotations perform best at 48.9%. For kNN A to B, Head-only, position and 6D rotations is best at 31.1%, and no condition performed above 25% accuracy for

Velocity B → A RF	**Position + Euler**	**Position + Quaternion**	**Position + SixD**	**Position**	**Euler**	**Quaternion**	**SixD**
Head + DomH + OffH	42.2%	44.4%	40.0%	33.3%	31.1%	40.0%	40.0%
Head + DomH	37.8%	35.6%	42.2%	28.9%	26.7%	28.9%	31.1%
Head + OffH	26.7%	35.6%	35.6%	17.8%	15.6%	35.6%	40.0%
DomH + OffH	31.1%	37.8%	28.9%	17.8%	22.2%	28.9%	26.7%
Head	28.9%	35.6%	46.7%	17.8%	17.8%	26.7%	33.3%
DomH	28.9%	26.7%	26.7%	13.3%	15.6%	20.0%	17.8%
OffH	22.2%	26.7%	24.4%	11.1%	11.1%	17.8%	20.0%

Table 11: Between-session identification accuracy for Random Forest, with velocity and angular velocity data, trained on data from B and evaluated with data from session A.

B to A.

8 Discussion

8.1 Tracking data from an unspecified task is likely not sufficient for identifying people against their will with these machine learning models

Our findings generally show that across the board, it's feasible to attain impressive results for identifying participants within the same task. The results of our analyses also demonstrate, however, that even when models are provided with substantially more data to evaluate over, they still fail to identify participants at nearly the same level of accuracy when attempting to evaluate over a slightly different task. This result is important because it demonstrates that XR practitioners hoping to develop motion-based authentication should present users with the same task for encoding and authentication. Likewise, those hoping to de-anonymize XR usage data should, if possible, train their models on identified samples of users performing parts of the tasks in their target data set, because even similar tasks yield between sessions yield worse results.

Furthermore, it is worth noting that in an ecologically valid between-task identification scenario, where one knew the labels for session A, but not B, one might attempt to use the best-performing classifier in their labelled training data (i.e., all three trackers with position and quaternion rotations, which performed at 96% accuracy). This condition applied between sessions yields relatively poor performance, however, at only 48.9% accuracy.

8.2 More tracked devices yields better identification accuracy within the same task

As shown in Tables 4 and 5, across all position and orientation conditions, the inclusion of additional trackers yielded better accuracies. This suggests that within a given task and session, none of the tracked objects contributed data that caused the models to overfit to the training subset. Perhaps unsurprisingly, we also note that of the 2-device and 1-device conditions, those that included the head performed better than those without, when position was included. This is aligned with the results of the ablation results presented by [7], in which removal of the features related to the head Y-value resulted in the greatest drop in identification accuracy.

8.3 More tracked devices does not yield better identification accuracy across similar but slightly different tasks

As shown in Tables 6 and 7, the inclusion of trackers other than the head yielded worse accuracies, suggesting that they contributed to noise in the training of the model. We were focused on the ecologically valid potential privacy issue in which a malicious actor has a sample of labeled, trained data. Because of our study design, we are unfortunately unable to separate out if this may have been due in part to the model encoding physiological features of the user such as their height, as opposed to features related to their environment as indicated by [16].

8.4 Using position and orientation generally yields higher identification accuracy

Across both the within- and between-session conditions, the inclusion of both position and orientation generally yielded higher identification accuracy than using exclusively one or the other. These results indicate that broadly when including the data from a given set of trackers, it appears to be helpful to make use of both the position and orientation tracking, if afforded by the system.

Interestingly, we find that in many of our conditions, the six-dimensional representation proposed by Zhou et al. [19] not to perform significantly better. We posit that this may be partially a result of the manner in which we aggregate positions. Because each per-second feature vector was considered in isolation, moments where the quaternion or Euler values would have discontinuities was outnumbered by the samples of data without such discontinuities. It is possible that this representation may be more useful for a more time-dependent approach such as LSTM.

8.5 Identities can be partially obfuscated by encoding the velocities of tracked devices instead of positions

While one might expect that using feature vectors generated from the first-order derivative to potentially encode data more specific to users and yield classifiers that are better capable of generalizing across tasks, we found that between-task performance was generally worse with the velocity-based features than with the positional ones, diminishing from 82.2% to less than 50% for Random Forest. While this is still much more than the random likelihood of a correct guess, this diminishing suggests that while velocity-based features may have some use for identification, they may need a specific featurization outside of the scope of this paper to yield accuracies on par with our positional data.

8.6 Limitations

In this work, we choose not to optimize some of the hyperparameters that are intrinsic to the models explored (such as the depth of trees in Random Forest, or k in kNN). This decision is due in part to the scope of the explored hyperparameters in this work requiring the evaluation of 1176 $(2*4*7*7*3)$ models, resulting in further hyperparameter exploration to be untenable. Additionally, our approach categorized each feature vector in isolation, not taking advantage of the temporal nature of our data.

Another limitation is that we examined the data from two similar but distinct tasks. It's possible that our results could be different with different builds or with fundamentally different tasks. We believe that this is a somewhat representative example of identifiability between similar tasks in which the user is performing a task that is prescriptively defined as a sequence of steps, but varying results may be found for environments in which the interactive objects are different, appear in different locations, or lack the instructional context our application contains.

Further, while our total set of data may appear large (1.7 million samples), this represents the data of 45 users and a total of a little more than 5 hours of time. If we were to make use of Deep Learning methods on this amount of data, it would be likely to overfit to the individuals. While this amount of data is less than some previous work like that by Miller et al. [7] or Moore et al. [11], it is more than several of the other works exploring VR identifiability, and will be made available as a resource for future research.

One final limitation to address is the arguable presence of a confound due to differences in the builds in addition to the participants doffing and donning the VR equipment. Some existing work, such as that by Asish et al. [1] makes use of multiple scenes and contexts without removal of the headset. With the between-session portion of this work, our focus was on identifiability between two individual VR sessions. Because our context is in an ecologically valid training scenario, we anticipated learning affects to moderate participants' motions through the instructions. This would still result in unique behavior from session to session, so we opted for two unique, but slightly different builds.

8.7 Future Work

This work demonstrates that training a classifier within a session of data can overfit to features specific to the task a participant is experiencing, and that using this data from one unspecified task may be insufficient to identify participants in other tasks against their will. In our dataset, participants were exposed to one session in VR, exited VR, then re-entered for the other task. Exploring multiple tasks within a single VR session, and alternatively the same task, longitudinally across multiple sessions will be an important next step for improving our understanding of the variability in tracking data across experiences and sessions. A further logical next step for this work would be to design additional experiences, and examine to what degree training classifiers on samples of data from multiple distinct tasks can improve accuracy in identifying people in unseen tasks.

9 Conclusion

In this paper, we explored the identifiability of participants when presented with two distinct but similar tasks in VR. We first looked at identifiability within sessions and we examined the inclusion and exclusion of data across several conditions by tracked object as well as the kind of data contained therein. Our search over this space of data should help form an understanding of what is contributing to accuracy and what is overfitting for different training and testing conditions. Further, by examining temporally close data from two VR sessions, we hoped to expand our understanding of identifiability as it relates to variables specific to a given session. While we found that within a VR session, increasing the sources of data generally improved the accuracy of the models identifications, we found some sets of data to be contributing data that overfits to the session, thus resulting in worse accuracy. Finally, by exploring a velocity representation, we find different sets of data to be useful for identification, warranting further exploration.

References

[1] S. M. Asish, A. K. Kulshreshth, and C. W. Borst. User identification utilizing minimal eye-gaze features in virtual reality applications. In *Virtual Worlds*, vol. 1, pp. 42–61. MDPI, 2022.

[2] B. David-John, D. Hosfelt, K. Butler, and E. Jain. A privacy-preserving approach to streaming eye-tracking data. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2555–2565, 2021. doi: 10.1109/TVCG.2021.3067787

[3] X. Hu, A. G. Moore, J. Coleman Eubanks, A. Aiyaz, and R. P. McMahan. Evaluating interaction cue purpose and timing for learning and retaining virtual reality training. In *Symposium on Spatial User Interaction*, pp. 1–9, 2020.

[4] G. Lee, Z. Deng, S. Ma, T. Shiratori, S. S. Srinivasa, and Y. Sheikh. Talking with hands 16.2 m: A large-scale dataset of synchronized body-finger motion and audio for conversational motion analysis and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 763–772, 2019.

[5] 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*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445528

[6] J. Liebers, P. Horn, C. Burschik, U. Gruenefeld, and S. Schneegass. Using gaze behavior and head orientation for implicit identification in virtual reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–9, 2021.

[7] M. R. Miller, F. Herrera, H. Jun, J. A. Landay, and J. N. Bailenson. Personal identifiability of user tracking data during observation of 360-degree vr video. *Scientific Reports*, 10(1):1–10, 2020.

[8] A. G. Moore, R. P. McMahan, H. Dong, and N. Ruozzi. Extracting velocity-based user-tracking features to predict learning gains in a virtual reality training application. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 881–890. IEEE Computer Society, Los Alamitos, CA, USA, nov 2020. doi: 10.1109/ISMAR50242.2020.00099

[9] A. G. Moore, R. P. McMahan, H. Dong, and N. Ruozzi. Personal identifiability and obfuscation of user tracking data from vr training sessions. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 221–228, 2021. doi: 10.1109/ISMAR52148.2021.00037

[10] A. G. Moore, R. P. McMahan, H. Dong, and N. Ruozzi. Personal identifiability of user tracking data during vr training. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 556–557. IEEE, 2021.

[11] A. G. Moore, R. P. McMahan, and N. Ruozzi. Exploration of feature representations for predicting learning and retention outcomes in a vr training scenario. *Big Data and Cognitive Computing*, 5(3):29, 2021.

[12] 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*, IWSPA '18, p. 23–30. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3180445.3180450

[13] I. Olade, C. Fleming, and H.-N. Liang. Biomove: Biometric user identification from human kinesiological movements for virtual reality systems. *Sensors*, 20(10), 2020. doi: 10.3390/s20102944

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[15] K. Pfeuffer, M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt. 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*, CHI '19, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300340

[16] C. Schell, A. Hotho, and M. E. Latoschik. Comparison of data representations and machine learning architectures for user identification on arbitrary motion sequences. *arXiv preprint arXiv:2210.00527*, 2022.

[17] P. P. Tricomi, F. Nenna, L. Pajola, M. Conti, and L. Gamberini. You can't hide behind your headset: User profiling in augmented and virtual reality. *arXiv preprint arXiv:2209.10849*, 2022.

[18] B. Xie, H. Liu, R. Alghofaili, Y. Zhang, Y. Jiang, F. D. Lobo, C. Li, W. Li, H. Huang, M. Akdere, et al. A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, 2:645153, 2021.

[19] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5745–5753, 2019.