Inferring Private Personal Attributes of Virtual Reality Users from Ecologically Valid Head and Hand Motion Data

Vivek Nair[1] UC Berkeley
Christian Rack[2] University of Würzburg
Wenbo Guo[3] Purdue University
Rui Wang[4] Carnegie Mellon
Shuixian Li[5] UC Berkeley
Brandon Huang[6] UC Berkeley
Atticus Cull[7] UC Berkeley
James F. O'Brien[8] UC Berkeley
Marc Latoschik[9] University of Würzburg
Louis Rosenberg[10] Unanimous AI
Dawn Song[11] UC Berkeley

ABSTRACT

Motion tracking "telemetry" data lies at the core of nearly all modern virtual reality (VR) and metaverse experiences. While generally presumed innocuous, recent studies have demonstrated that motion data actually has the potential to uniquely identify VR users. In this study, we go a step further, showing that a variety of private user information can be inferred just by analyzing motion data recorded from VR devices. We conducted a large-scale survey of VR users (N=1,006) with dozens of questions ranging from background and demographics to behavioral patterns and health information. We then obtained VR motion samples of each user playing the game "Beat Saber," and attempted to infer their survey responses using just their head and hand motion patterns. Using simple machine learning models, over 40 personal attributes could be accurately and consistently inferred from VR motion data alone. Despite this significant observed leakage, there remains limited awareness of the privacy implications of VR motion data, highlighting the pressing need for privacy-preserving mechanisms in multi-user VR applications.

Index Terms: Security and privacy; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Computing methodologies—Machine learning

1 INTRODUCTION

With the recent emergence of affordable standalone virtual reality (VR) devices like the Meta Quest 2, VR technology has begun to reach mass-market adoption for the first time, with nearly 10 million VR systems sold just in 2022 [14]. While the major proponents of VR envision the ultimate use of these devices to create an immersive virtual "metaverse" where users meet to work, learn, and socialize, contemporary adoption of VR has been driven primarily by gaming. As of early 2023, VR games, including "Beat Saber," constitute 91 of the 100 most popular VR applications [26].

On conventional platforms, gaming is typically perceived as amongst the most innocuous classes of applications from a security and privacy perspective, while, for example, social media, receives far more attention in this regard. However, recent research indicates that the same may not be true in VR. Researchers have already demonstrated the ability to uniquely identify users in VR [11, 17] and construct adversarial VR applications that harvest a variety of private user data while being disguised as harmless games [16].

[1]e-mail: vcn@berkeley.edu
[2]e-mail: christian.rack@uni-wuerzburg.de
[3]e-mail: henrygwb@purdue.edu
[4]e-mail: ruiwang3@andrew.cmu.edu
[5]e-mail: shuixian.li@berkeley.edu
[6]e-mail: zhaobin@berkeley.edu
[7]e-mail: atticuscull@berkeley.edu
[8]e-mail: job@berkeley.edu
[9]e-mail: marc.latoschik@uni-wuerzburg.de
[10]e-mail: louis@unanimous.ai
[11]e-mail: dawnsong@berkeley.edu

While the prospect of malicious VR applications poses a legitimate security and privacy threat, most VR games are not deliberately designed to harvest user information. A typical VR game does not include the adversarial challenge and response mechanisms designed to reveal user data in prior work, but does often still broadcast motion data to other players in order to facilitate multi-player functionality. In this study, we aim to explore the extent to which popular non-adversarial VR games may inadvertently leak private information about their users by revealing their head and hand motion patterns.

It has long been understood that individuals exhibit distinct motion patterns as determined by their unique physiology and muscle memory. Since as early as 1977, studies have demonstrated that these motion patterns can be used not only to uniquely identify individuals [5], but also to infer personal characteristics such as age [8] and gender [10, 19]. However, the extent to which these findings are applicable to the motion data observable in VR is not yet well understood; although full-body tracking systems are on the horizon, most VR devices today only collect head and hand motion.

To determine whether private information can be inferred from the head and hand motion data broadcast by a typical multi-player VR game, we surveyed players of the popular VR rhythm game "Beat Saber." Players were asked a series of about 50 questions, ranging from demographics like age and gender to personal background, behavioral patterns, and health information. Additionally, participants were asked to provide links to motion-capture recordings of themselves playing Beat Saber on their own personal VR devices.

After collecting data attributes and motion samples from over 1,000 users, we designed 50 binary classification problems based on thresholding the dataset (e.g., "old" vs "young," or "rich" vs "poor"). We then trained and tested a deep-learning binary classifier that ingested a sequence of motion data and produced a binary classification for each attribute. We found that over 40 of the 50 attributes could consistently and reliably be inferred from user motion data alone. Thus, while these users may hold the presumption of anonymity in a VR gaming setting, this presumption is evidently flawed. Not only are their movement patterns revealing their identity [11, 17], our results imply these patterns could actually be exposing a plethora of information about them to the device, application, server, and even other users within the same virtual environment.

The goal of our work is not to provide an optimal approach for inferring any particular attribute from VR motion data. Rather, we aim to demonstrate, with high statistical significance, that a wide variety of personal and privacy-sensitive variables can be inferred from head and hand motion, in order to highlight the urgent need for privacy-preserving mechanisms in multi-user VR applications.

Contributions:

- We surveyed over 1,000 VR users to generate a comprehensive dataset of motion recordings and user data attributes (§3).
- We present a general-purpose machine learning architecture for inferring user data from head and hand motion streams (§4).
- We demonstrate over 40 classes relating to personal user data can be inferred from motion data in standard VR games (§5).

DOI 10.1109/VRW62533.2024.00094

2 Background and Related Work

This work follows the 2023 Garrido et al. VR privacy systematization of knowledge (SoK) [7], which provides a standard model of VR information flow and threat actors for VR security and privacy research. In this section, we will summarize the threat model of Garrido et al., as well as related work in this area, so as to position our study within the broader landscape of VR privacy research.

2.1 VR Information Flow and Threat Model

Most modern consumer-grade VR systems include at least a head-mounted display (HMD) and two hand-held controllers.[1] These systems use an array of sensors to measure the position and orientation of tracked components in 3D space, providing six degrees of freedom (6DoF) per tracked object, typically captured between 60 and 144 times per second, resulting in a "telemetry stream."

Many advanced VR devices now contain additional sensors, such as LIDAR arrays, microphones, cameras, full-body tracking, and eye-tracking systems. However, in this paper, we focus exclusively on the basic telemetry data stream consisting of head and hand motion. Thus, any information that can be derived from this data stream should be applicable to nearly all VR devices and users. We expect that the privacy threats discussed in this paper would be exacerbated when motion data is combined with other data streams.

In a typical VR application, the telemetry data stream is forwarded to various parties via the following standard information flow:

1. *Device*. Telemetry data originates at the VR device, which measures tracked objects with 6DoF at between 60 Hz and 144 Hz.
2. *Application*. A client-side application uses the telemetry data to render a series of auditory, visual, and haptic stimuli, creating an immersive 3D virtual experience for the end user.
3. *Server*. In the case of a multiplayer or metaverse experience, the application forwards the telemetry stream to an external server.
4. *Other Users*. The server, in turn, forwards this data to other users, such that a virtual representation (or "avatar") of each user can be rendered on the devices of other users in the same virtual world.

Because each entity in the above information flow has access to the telemetry stream of a target user, and could theoretically use it to make adversarial inferences, they are each considered potential adversaries in the Garrido et al. threat model. Adversaries in this model exist on a continuum, as shown in Figure 1, with adversaries becoming "weaker" from left to right due to potential interference, such as compression or transformation, at each step of the data flow.

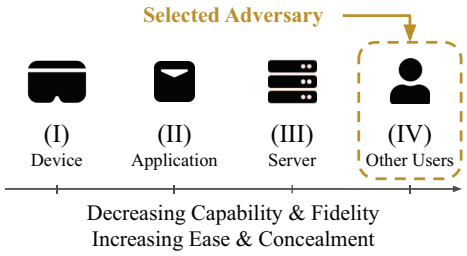

Figure 1: VR threat continuum and selected adversary for this work.

In this paper, we have chosen to focus on the user adversary (IV) by using only motion data that would normally be available to ordinary users of a multi-user VR application. Because this is considered the weakest adversary in the Garrido et al. threat model, attacks available to this user can usually also be performed by all other adversaries in the system, while also being amongst the hardest attacks to detect due to their remote and decentralized nature.

[1] Camera-based hand tracking is an increasingly common alternative.

2.2 Human Motion Biometrics

Since at least the 1970s, researchers have known that individuals reveal identifiable characteristics via their motion. In a 1977 study, Cutting and Kozlowski demonstrated that the gender of 6 individuals could be inferred by a panel of participants using the motion of 8 tracked objects affixed to the body [10]. The study took place before the advent of modern computer graphics, so the authors creatively resorted to taping highly reflective objects to a number of points on the participants' bodies. The researchers then streamed a camera feed of the subjects through a CRT television monitor, and increased the contrast until the participants' silhouettes disappeared and only the individual points of light could be seen. Using only the moving points of light visible on the screen, participants were able to infer the gender of the original subjects with up to 63% accuracy ($p < 0.05$).

More recently, Pollick et al. (2005) [19] used statistical features to achieve 79% accurate identification of gender from motion, while Sarangi et al. (2020) [24], and others have used machine learning to achieve inference accuracies of 83% or more. Beyond gender, Jain et al. (2016) [8] found that the motion of children can be differentiated from that of adults with 66% accuracy. Overall, researchers have long known that human motion patterns can be used to infer a variety of personal attributes from human subjects in a laboratory setting.

In one sense, by deriving data from the motion of human body parts tracked in 3D space, these results use data that is highly similar to the motion captured by a VR device. On the other hand, basic VR devices capture only 3 tracked objects, rather than the 8 or more used for motion capture in laboratory studies. Thus, it is not clear whether these previous findings will be transferable to the VR setting.

2.3 VR Privacy Attacks

In addition to proposing a standard information flow and threat model, the Garrido et al. SoK tracks a number of existing works in the VR privacy domain. The majority of these works are categorized as "identification," in which VR users are deanonymized or tracked across sessions based on their movement patterns. For instance, Miller et al. (2020) [11] performed a lab study of 511 users, and then correctly identified users within the pool of 511 with 95% accuracy using a random forest model. In the largest study to date, Nair et al. (2023) [17] used a LightGBM model to identify over 55,000 Beat Saber players with 94.3% accuracy from their motion. Several similar studies have also been produced [7, 12, 13].

A relatively smaller portion of the existing work is categorized as "profiling," whereby specific attributes, such as age or gender, are inferred from users in VR. In one such study, Tricomi et al. (2023) [27] accurately infer the gender and age of about 35 VR users, using eye tracking data in addition to head and hand motion.

The remaining studies in this field have focused on the adversarial design of malicious VR applications. Specifically, Aliman and Kester (2020) [2] evaluate the risk of generative AI actors harvesting user data in VR, while Nair et al. (2022) [16] present "MetaData," an VR escape room game designed to trick users into revealing personal information via covert adversarially-designed puzzles. In the case of MetaData, the authors were able to recover over 25 personal data attributes, from anthropometrics like height and wingspan to demographics like age and gender, within just a few minutes of gameplay. However, sensors like microphones and cameras were used in addition to motion data to make these inferences.

While the existing literature strongly indicates that VR applications can be used for profiling, it does so mostly by using applications that are explicitly adversarial in design, or by using signals beyond the basic motion data widely available in VR. By contrast, this work aims to demonstrate that profiling is possible even by the weakest known class of adversaries, in popular benign applications like Beat Saber, and from head and hand motion data alone.

3 Data Collection

We chose to use "Beat Saber" as the model application and main data source for this study, primarily due to the popularity of the game and the relative availability of user data from this application. Furthermore, to maximize ecological validity, we chose to only include participants who were already Beat Saber players. In this section, we provide background information about Beat Saber and describe our methods for collecting data from existing players.

3.1 Beat Saber

"Beat Saber" [6] is a VR rhythm game where players slice blocks representing musical beats with a pair of sabers, one held in each hand. With over 6 million copies sold, Beat Saber is the highest-grossing VR application of all time [29], and is a representative example of a non-adversarial VR game with multi-player functionality.

Figure 2: Dynamic objects in "Beat Saber."

Beat Saber contains a number of "maps," which consist of an audio track as well as a series of objects presented to the user in time with the audio. These objects include "blocks," which the player must hit at the correct angle with the correct color saber, "bombs," which the player must avoid hitting with their sabers, and "walls," which the player must avoid with their head (see Fig. 2). At the end of the map, the player is awarded a number of points based on their level of accuracy in completing these tasks. Hundreds of official maps have been added to the game, and over 100,000 user-created maps can be played by installing open-source game modifications.

3.2 BeatLeader

"BeatLeader" [21] is a popular open-source Beat Saber extension and website that offers a third-party leaderboard system for user-created Beat Saber maps. Beat Saber enthusiasts may choose to install the extension in order to compete with other players to achieve a higher "rank" on the leaderboards for popular maps. After playing a Beat Saber map with the BeatLeader extension enabled, scores are automatically uploaded to a globally-visible leaderboard, with over 4 million scores having been uploaded to the platform to date.

When uploading a score to BeatLeader, a recording of the user's motion telemetry during play is automatically captured and attached to their submission. The recording, which utilizes the "Beat Saber Open Replay" (BSOR) [22] format, is then made publicly available on the BeatLeader website so that it can be used to verify the authenticity of the submitted score.

3.3 Survey Procedures

We partnered with the administrators of BeatLeader to conduct an official survey of BeadLeader users, consisting of about 50 personal questions across 9 categories of information. Beat Saber players were invited to take the survey via an announcement released through the official Twitter and Discord accounts of BeatLeader. Participation in the survey was voluntary, with all questions being optional, and no consequences for non-participation. Participants were not monetarily compensated but were given the option to add a unique badge to their BeatLeader profile in recognition of their contribution.

The survey was conducted from April 15th, 2023 to May 1st, 2023, with 1,006 responses collected in that time. The categories of information collected were as follows:

1. *Participation*. Participants provided links to their BeatLeader profile from which motion capture recordings could be obtained.
2. *Demographics*. Participants were asked a variety of demographic questions based on the 2020 United States Census [3].
3. *Specifications*. Participants shared an automatically-generated system report containing various computer specifications.
4. *Background*. Participants were asked about their past history with musical instruments, rhythm games, dancing, and athletics.
5. *Health*. Participants were asked about their mental and physical health status and disabilities as well as their visual acuity.
6. *Habits*. Participants were asked about their habits relating to Beat Saber, such as their warmup routine prior to playing.
7. *Environment*. Participants were asked about the sizes and locations of the areas in which they typically play Beat Saber.
8. *Anthropometrics*. Participants were asked to measure various physical dimensions of their body, such as height and wingspan.
9. *Clothing*. Participants were asked about the clothing and footwear they typically wear while playing Beat Saber.

The exact questions asked in each section are given in Appendix B.

3.4 Motion Recordings

During the informed consent procedure for the survey, participants also gave us permission to use the publicly-available motion capture recordings from their BeatLeader profile to infer the attributes contained in their survey responses. Accordingly, we obtained the head and hand motion recordings of each participant from the BOXRR-23 dataset [18] using the profile indicated in their survey responses. For participants with more than 100 recordings in the BOXRR-23 dataset, only the latest 100 recordings were utilized.

3.5 Ethical Considerations

We conducted this project with significant attention to ethical considerations. Specifically, we refrained from asking questions that could be viewed as overly sensitive, and did not solicit responses from vulnerable populations, including minors under the age of 18. Participants were required to read and agree to a thorough informed consent document prior to inclusion in the study.

An additional source of data was scoring information collected by BeatLeader. This data was already widely publicly available prior to this study, and was available for research in accordance with BeatLeader's privacy policy. Further, participants explicitly agreed to our specific use of this data via the informed consent process.

Participants were not monetarily compensated or given anything of substantial value for their participation in the survey, nor penalized for non-participation. Every question in the survey was optional. Thus, participants were never unduly pressured to provide information that they were uncomfortable with disclosing.

Because the survey responses include sensitive information, such as health status, we followed the strictest data handling standards and guidelines offered by our institution throughout this study.

Overall, we believe this research constitutes a net benefit to society by highlighting the magnitude of the VR privacy threat and motivating future work on defensive countermeasures.

This study has been reviewed and approved by UC Berkeley's Institutional Review Board (IRB) as protocol #2023-03-16120.

4 METHOD

In this section, we describe our method for determining which of the survey responses are inferrable from VR telemetry data. Specifically, we describe a machine learning model architecture that attempts to infer user data attributes based on a sequential input containing their head and hand motion. Importantly, our goal in this section is not to describe an optimal architecture for inferring any particular attribute, such as age or gender, from motion data. Rather, we aim to describe a general-purpose method for producing binary classifications from VR motion data, and use this method to determine which attributes are present in the motion data with high statistical significance.

4.1 Binary Classifications

Our survey results contain a variety of attribute types, including categorical variables such as ethnicity or languages spoken, and numerical variables like height or age, all with different observed distributions. We began by choosing 50 attributes that we speculated had a reasonable chance of being inferred from motion patterns. To simplify our analysis, we then turned each of these attributes into a binary classification problem. For example, marital status was turned into a binary classification of "never married" versus all other responses (married, divorced, etc.). For continuous attributes, such as height, a wide rejection band was usually incorporated. The exact binary splits for all attributes are given in Appendix B.

Using this approach allows us to use a single binary classification model architecture and statistical analysis technique for all attributes being considered. This simplified approach is sufficient for our purposes of demonstrating whether the attribute can be inferred from VR motion data, though regression or multi-class classification models may be more suitable for use in a real-world deployment.[2]

4.2 Model Architecture

We evaluated the efficacy of a variety of machine learning architectures, including Random Forest, CNN, LSTM, and Transformer models, for performing our binary classification task using the sequential motion data. We found the Transformer-based models to be most effective at inferring a majority of the chosen attributes.

The Transformer model [28] incorporates a self-attention mechanism to capture dependencies and relationships within an input sequence. Unlike other deep learning models that process the elements of a sequence sequentially, the Transformer simultaneously processes all elements in parallel, allowing it to weigh the importance of each element in the context of the whole sequence.

FIG. 3 illustrates our Transformer implementation. Input sequences first go through a projection layer that prepares the features by increasing the dimensionality to the embedding size of the Transformer. Following this, the data pass an encoding layer that applies a sinusoidal positional encoding, which adds information about the relative position of each element in the input sequence. This step is important, as Transformers do not have an inherent notion of order or position. Next, we use the Transformer encoder component to generate a contextualized representation of the input sequence. Finally, a fully connected output layer reduces the encoder output to a scalar value, which provides the binary classification.

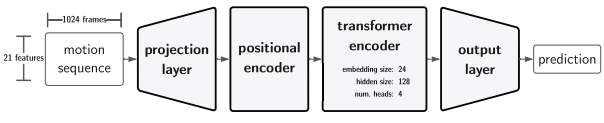

Figure 3: Transformer model architecture.

[2]We stress that deployment in a context where the user has not specifically and knowingly agreed to this type of monitoring would raise significant ethical concerns, particularly if the data remained linked to the user's identity.

4.3 Model Input

An advantage of transformer models is that they are intrinsically well-suited for handling time-series data. We thus chose to use a sequential featurization method to encode the motion of VR users in 3D space over a period of time. At a given time step ("frame"), we capture the position and orientation of three objects (the user's head and two hands) in 3D space. Each tracked object is captured using three positional coordinates and four orientation coordinates expressed as a quaternion. In total, 7 coordinates are taken for each object, resulting in a total of 21 values captured per frame. For each motion recording, we sample the first 1,024 frames to provide as input to our model. Thus, any given recording is represented by a (21×1024)-dimensional input; recordings with less than 1024 frames were zero-padded. The frames were sampled at their original frame rate without interpolation or normalization, as the model's normalization layer already allows it to rescale inputs internally.

4.4 Data Split

Using the BeatLeader database, we downloaded the 100 most recent motion recordings from each of the users who responded to our survey. We then converted each recording into a (21×1024)-dimensional input using the featurization method of §4.3.

For each of the 50 attributes, we selected 20 users from each of the two classes to split between testing and validation sets, with the remaining users being used for training. We then resampled the training sequences to select $10,000$ recordings for training each class. As such, all three sets were perfectly balanced between both classes of every binary attribute, with $10,000$ recordings for training each class and $1,000$ recordings for validating and testing each class. This process was repeated across 3 to 7 Monte Carlo cross-validations [30] for each attribute to assess statistical significance.

4.5 Training

We evaluated the machine learning approach by using PyTorch v2.0.1 to train and test one binary classification model for each of the 50 selected attributes. We utilized the Adam optimization algorithm [9] with a binary cross-entropy (BCE) loss function. Each model was trained across 100 epochs, with the best-performing epoch then being selected using a validation set. The evaluation was performed using a single machine with an AMD Ryzen 9 5950X 16-core CPU, 128 GB of DDR4 RAM, and an NVIDIA GeForce RTX 3090 GPU. With this setup, each model took an average of 37 minutes to train and test, with the entire evaluation taking approximately 31 hours.

4.6 Evaluation Metrics

Because our sampling technique always includes the same number of sequences and users in each class, the statistical significance of these results can be evaluated using a cumulative binomial test where n is the number of samples, K is the number of correct predictions, and $p_\emptyset$ is 0.5. We use this as our primary target metric in §5. The use of completely balanced training, testing, and validation sets substantially diminishes the need for more nuanced statistical tests, such as the F_1-score [25] or Cohen's kappa [4].

4.7 Hyperparameter Tuning

We performed a tuning sweep of the relevant hyperparameters (hidden size, learning rate, etc.) using just two attributes, `StandaloneGrip` and `Sex`. The hyperparameters that maximized the significance of these attributes per the metrics in §4.6 were then used throughout our evaluation and are provided in Appendix A.

	Per Sequence				Per User			
Attribute	**Total #**	**Test #**	**Accuracy**	**Significance**	**Total #**	**Test #**	**Accuracy**	**Significance**
StandaloneGrip	31,100	6,000	85.9%	p <0.001	311	60	91.7%	p <0.001
Height	19,100	6,000	76.5%	p <0.001	191	60	86.7%	p <0.001
Controller	33,200	6,000	81.2%	p <0.001	332	60	85.0%	p <0.001
Weight	9,800	6,000	73.6%	p <0.001	98	60	85.0%	p <0.001
FootSize	9,100	6,000	73.2%	p <0.001	91	60	85.0%	p <0.001
Country	33,300	6,000	60.3%	p <0.001	333	60	81.7%	p <0.001
RhythmGames	10,900	6,000	63.5%	p <0.001	109	60	80.0%	p <0.001
Age	62,300	6,000	64.9%	p <0.001	623	60	78.3%	p <0.001
TotalPlayTime	34,400	6,000	67.7%	p <0.001	344	60	78.3%	p <0.001
Headset	65,000	6,000	66.9%	p <0.001	650	60	76.7%	p <0.001
LeftArm	10,300	6,000	65.2%	p <0.001	103	60	76.7%	p <0.001
RightArm	10,200	6,000	64.9%	p <0.001	102	60	75.0%	p <0.001
Athletics	8,700	6,000	59.1%	p <0.001	87	60	75.0%	p <0.001
MaritalStatus	81,400	6,000	60.2%	p <0.001	814	60	73.3%	p <0.001
EmploymentStatus	64,200	6,000	65.1%	p <0.001	642	60	71.7%	p <0.001
AnyRhythmGames	83,000	6,000	54.8%	p <0.001	830	60	70.0%	p <0.001
Ethnicity	73,900	6,000	59.7%	p <0.001	739	60	70.0%	p <0.001
SteamComputerFormFactor	51,300	6,000	58.5%	p <0.001	513	60	70.0%	p <0.001
Footwear	36,700	6,000	60.5%	p <0.001	367	60	70.0%	p <0.001
AnyVRRhythmGames	83,000	8,000	56.8%	p <0.001	830	80	68.8%	p <0.001
Income	76,700	8,000	55.0%	p <0.001	767	80	68.8%	p <0.001
Wingspan	16,000	8,000	59.9%	p <0.001	160	80	68.8%	p <0.001
Handedness	71,600	10,000	55.2%	p <0.001	716	100	66.0%	p <0.001
HandLength	51,000	8,000	58.5%	p <0.001	510	80	66.3%	p = 0.002
SubstanceUse	69,200	10,000	55.9%	p <0.001	692	100	64.0%	p = 0.002
Preparation	39,400	8,000	58.2%	p <0.001	394	80	65.0%	p = 0.005
LowerBody	29,500	8,000	55.9%	p <0.001	295	80	65.0%	p = 0.005
Lenses	80,900	8,000	55.3%	p <0.001	809	80	65.0%	p = 0.005
Languages	80,700	8,000	56.5%	p <0.001	807	80	65.0%	p = 0.005
SteamOperatingSystemVersion	50,800	8,000	58.4%	p <0.001	508	80	65.0%	p = 0.005
Music	29,600	8,000	53.6%	p <0.001	296	80	65.0%	p = 0.005
AnyMentalDisabilities	83,000	10,000	52.6%	p <0.001	830	100	63.0%	p = 0.006
Sex	76,300	10,000	56.5%	p <0.001	763	100	63.0%	p = 0.006
AnyPhysicalDisabilities	83,000	10,000	54.5%	p <0.001	830	100	62.0%	p = 0.010
ReactionTime	9,800	14,000	53.1%	p <0.001	98	140	60.0%	p = 0.011
AnyMusic	83,000	8,000	55.7%	p <0.001	830	80	62.5%	p = 0.016
AnyAthletics	19,900	8,000	55.7%	p <0.001	199	80	61.3%	p = 0.016
EducationalStatus	62,200	8,000	57.1%	p <0.001	622	80	60.0%	p = 0.028
IPD	6,700	8,000	55.8%	p <0.001	67	80	60.0%	p = 0.028
Dance	82,000	10,000	52.3%	p <0.001	820	100	59.0%	p = 0.028
PoliticalOrientation	33,100	10,000	53.5%	p <0.001	331	100	58.0%	p = 0.044
UpperBody	47,200	10,000	52.0%	p <0.001	472	100	57.0%	p = 0.067
SteamProcessorLogicalCores	33,500	10,000	51.0%	p = 0.021	335	100	56.0%	p = 0.136
HadCOVID	83,000	10,000	54.4%	p <0.001	830	100	55.0%	p = 0.136
CaffinatedBeverages	40,800	10,000	52.9%	p <0.001	408	100	55.0%	p = 0.136
RoomArea	33,100	8,000	50.5%	p = 0.183	331	80	56.3%	p = 0.157
PhysicalFitness	7,800	12,000	54.2%	p <0.001	78	120	55.0%	p = 0.158
SteamProcessorCPUVendor	51,600	10,000	49.2%	p = 0.953	516	100	53.0%	p = 0.309
SteamLighthouses	5,500	8,000	48.6%	p = 0.993	55	80	52.5%	p = 0.369
ColorBlindness	79,800	10,000	50.4%	p = 0.227	798	100	52.0%	p = 0.382

Table 1: Accuracy of inferring 50 attributes from head and hand motion data, with statistical significance calculated via binomial tests.

5 Results

After training a model for each of the tested attributes, we first generated a classification for all of the 100 sequences per user for every user in the testing set. Next, we generated a prediction for each user by taking the average classification across their 100 sequences. Table 1 shows the accuracy of the results per sequence and per, along with the p-values corresponding to the metrics described in §4.6.

Overall, we observe that 33 of the 50 attributes were predicted with high statistical significance ($p < 0.01$), and another 8 of 50 with moderate statistical significance ($p < 0.05$) on a per-user basis. On a per-sequence basis, 45 out of 50 attributes were highly significant ($p < 0.01$), and one was moderately significant ($p < 0.05$). The difference in significance is largely accounted for by sample size; in total, 100 times more recordings were present than users.

5.1 Macro Significance

Given that we evaluated 50 attributes in this work, only a portion of which were inferred with significant accuracy, it remains to be demonstrated that the overall evaluation was statistically significant. To assess the overall significance of our result, we performed a secondary evaluation in which the trained models from our main evaluation were tested with randomly-generated fictitious inputs. We then performed a Wilcoxon signed-rank test to compare the distribution of classification accuracy values across the 50 attributes on these fictitious inputs with the distribution of true results in Table 1. We found $p < 0.0001$ on both a per-sequence and per-user basis, indicating a high overall statistical significance of our results.

6 Discussion

Though the threat of adversarial game design in VR [2, 16] remains salient, in this study, we have shown that even a seemingly harmless VR rhythm game reveals enough motion data to infer a wide variety of user characteristics. "Beat Saber" is not particularly conducive to data harvesting, with a simple ruleset and interaction model. For instance, there are no in-game characters to interact with, which could reveal even more information than we already observed.

In comparison with prior work, the setting evaluated in this paper represents a realistic and challenging threat scenario. Unlike laboratory studies, which take advantage of a controlled environment with homogeneous hardware and firmware, our data comes from real VR users around the world with a wide variety of devices and environments. We further limited our models to only use head and hand motion data, discarding other data modalities used in prior work, and used the weakest adversary class for our evaluations.

Despite the inherent difficulty of this dataset, a large number of personal attributes were accurately and consistently inferrable from the XR motion data alone. These attributes go beyond the obvious anthropometric measurements to include a surprising amount of information about the player's background, demographics, environment, habits, and even health. Many of these attributes, such as disability status, could be considered highly private information by end users. Others, like political orientation, have historically been used by social media platforms for targeted advertising and could similarly leave VR users open to targeted influence campaigns.

There are also a number of avenues adversaries could pursue to further improve VR profiling capabilities. Since VR motion patterns are now known to constitute uniquely identifiable biometrics [17], adversaries are not limited to collecting data from a single application. Rather, they could leverage the identifiability of VR users to track them across applications and usage sessions, building a rich user profile over time. These risks are further exacerbated with the introduction of data from additional sensors, such as microphones, cameras, LIDAR arrays, and eye and body tracking, all of which may provide data beyond the head and hand motion considered herein.

Despite using terms like "attack" and "adversary" throughout this paper, there is nothing inherently unlawful about collecting data from VR users, who may agree to such data harvesting in terms of service agreements. However, while users have developed a level of understanding about data collection on the web, most lack awareness of the breadth of personal information that can be extracted from even the simplest of VR experiences, such as a rhythm game.

As it stands, major VR device manufacturers have been observed selling hardware at a loss of up to $10 billion per year [20]. Given the deep roots of major metaverse players in the advertising industry, there could be at least a temptation to leverage existing sales channels to monetize the rich user data inferable from motion in VR. Our results indicate were they to act on this temptation, VR manufacturers and developers could reasonably harvest a vast array of user information from motion data collected in VR applications, including sensitive data about their age, gender, weight, ethnicity, income, substance use, disability status, and more.

6.1 Limitations

The motion recordings used in this study originate entirely from a single game. While "Beat Saber" is the most popular VR game to date [29], and is a representative example of a non-adversarial VR game, we cannot yet demonstrate that our findings will generalize to other types of VR applications. Furthermore, we chose to only survey existing Beat Saber players, and are unsure whether novice players, who would potentially demonstrate less consistent movement patterns, would be equally susceptible to these inferences.

We used the game recordings to infer a series of attributes that were self-reported via an online survey, and were thus subject to the biases typically associated with self-reported data. The participants in this survey were also not representative of the general population; for example, over 80% of respondents were male. However, the sample is fairly representative of the current VR user population [1]. The distribution of each attribute is given in Appendix B. Unbalanced distributions did not inflate the reported results, as each binary class was rebalanced prior to training and testing.

Finally, a portion of the reported findings may be the result of hidden correlations rather than direct inference. For example, it is likely that some attributes like employment or marital status are not directly observable from motion, but are correlated to age, which can be inferred from motion data. These correlations could apply generally to human motion, but may also represent sampling or response biases. Appendix C shows the correlation between each pair of attributes. Due to the difficulty of explaining the internal function of deep learning models, we cannot easily determine the mechanisms of causality associated with each result. However, we consider the potential to infer this data from VR users to be noteworthy and concerning, regardless of the cause.

6.2 Future Work

Our major motivation for conducting this study is to highlight the need for future research into defensive countermeasures for safeguarding user motion data in VR. Unfortunately, doing so is non-trivial, as sharing motion data is genuinely required for a variety of legitimate purposes in VR applications. Furthermore, motion patterns are largely subconscious and are often deeply ingrained in muscle memory, and are thus hard to intentionally obscure.

A first step towards defending against motion-based inference attacks is gaining a better understanding of how they work. While deep learning models are notoriously difficult to explain, we hope to see future work that uses advanced model explainability techniques to better understand the mechanisms underlying our results.

Researchers have already evaluated the use of differential privacy in VR [15], with moderate success at obscuring anthropometric attributes like height and wingspan. However, it is unclear whether this approach would be effective at defeating the sequential machine learning models used to derive more complex attributes in this paper.

Future work should evaluate the use of "corruption models" designed to obscure sensitive attributes embedded in VR motion data while minimally impacting legitimate application functionality. Generative adversarial networks (GAN) have already been shown to be effective at hiding attributes like gender from sequential data [23], and could likely be applied to VR telemetry data streams as well.

Another defense worth exploring is the use of trusted execution environments (TEEs) to provide auditability for metaverse servers that utilize telemetry data. TEEs like Intel's SGX could provide a hardware-based attestation mechanism that allows users to verify that servers are only using their motion data for legitimate purposes.

7 Conclusion

With major new products like the Apple Vision Pro on the horizon, extended reality technologies are on track to soon become a ubiquitous means of accessing the internet. For the foreseeable future, motion tracking "telemetry" data will remain at the core of nearly all extended reality and metaverse experiences. In this study, we demonstrated that even in non-adversarial applications, this data stream carries significant privacy implications for VR users.

Privacy risks are not unique to VR, and users have grown accustomed to some level of data harvesting on most internet platforms. However, unlike in conventional web applications, users are largely unaware of the unique privacy risks associated with VR applications, and lack the suite of defensive tools, such as privacy-preserving browser extensions, that have been developed over time for the web.

Thus, we are currently at a crossroads. If nothing is done to improve the present security and privacy posture of virtual reality, it is poised to inherit an exaggerated version of the privacy issues that are prevalent on the web. However, we can also take the opportunity to learn from the history of browser-based attacks and defenses. We hope the results of this study motivate security and privacy practitioners to prioritize research in this field and build privacy-preserving mechanisms into the fabric of future metaverse platforms.

Acknowledgments

We appreciate the help of Allen Yang, Beni Issler, Bjoern Hartmann, Charles Dove, Ines Bouissou, Jason Sun, Julien Piet, and Xiaoyuan Liu. This work was supported by the Minderoo Foundation, the National Science Foundation, the National Physical Science Consortium, the Fannie and John Hertz Foundation, and the Berkeley Center for Responsible, Decentralized Intelligence.

References

[1] Report: Vive Users Are 95 Percent Male And Spend 5 Hours Per Week in VR, Feb. 2017.

[2] N.-M. Aliman and L. Kester. Malicious design in aivr, falsehood and cybersecurity-oriented immersive defenses. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, pp. 130–137, 2020. doi: 10.1109/AIVR50618.2020.00031

[3] U. C. Bureau. 2020 Census Results. Section: Government.

[4] J. Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46, 1960.

[5] J. E. Cutting and L. T. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9(5):353–356, May 1977. doi: 10.3758/BF03337021

[6] B. Games. Beat Saber. `https://beatsaber.com/`.

[7] G. M. Garrido, V. Nair, and D. Song. SoK: Data Privacy in Virtual Reality, Jan. 2023. arXiv:2301.05940 [cs].

[8] E. Jain, L. Anthony, A. Aloba, A. Castonguay, I. Cuba, A. Shaw, and J. Woodward. Is the Motion of a Child Perceivably Different from the Motion of an Adult? *ACM Transactions on Applied Perception*, 13(4):1–17, July 2016. doi: 10.1145/2947616

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization.

[10] L. T. Kozlowski and J. E. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics*, 21(6):575–580, Nov. 1977. doi: 10.3758/BF03198740

[11] 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*, 10:17404, Oct. 2020. Nature Publishing Group. doi: 10.1038/s41598-020-74486-y

[12] A. G. Moore, T. D. Do, N. Ruozzi, and R. P. McMahan. Identifying virtual reality users across domain-specific tasks: A systematic investigation of tracked features for assembly. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 396–404.

[13] 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.

[14] Q. a.-P. a. P. W. Movement. VR Headset Sales Underperform Expectations, What Does It Mean For The Metaverse In 2023?

[15] V. Nair, G. M. Garrido, and D. Song. Going incognito in the metaverse.

[16] V. Nair, G. M. Garrido, and D. Song. Exploring the unprecedented privacy risks of the metaverse, 2022.

[17] V. Nair, W. Guo, J. Mattern, R. Wang, J. F. O'Brien, L. Rosenberg, and D. Song. Unique identification of 50,000+ virtual reality users from head & hand motion data, 2023.

[18] V. Nair, W. Guo, R. Wang, J. F. O'Brien, L. Rosenberg, and D. Song. BOXRR-23: 4.7 million motion capture recordings from 105,852 extended reality device users, 2023.

[19] F. E. Pollick, J. W. Kay, K. Heim, and R. Stringer. Gender recognition from point-light walkers. *Journal of Experimental Psychology: Human Perception and Performance*, 31:1247–1265, 2005. American Psychological Association. doi: 10.1037/0096-1523.31.6.1247

[20] A. D. published. Oculus will sell you a Quest 2 headset that doesn't need Facebook for an extra $500. *PC Gamer*, Apr. 2021.

[21] V. Radulov. BeatLeader.

[22] V. Radulov. Beat Saber Open Replay, Sept. 2022.

[23] A. Rezaei, C. Xiao, J. Gao, and B. Li. Protecting sensitive attributes via generative adversarial networks, 12 2018.

[24] V. Sarangi, A. Pelah, W. E. Hahn, and E. Barenholtz. Gender perception from gait: A comparison between biological, biomimetic and non-biomimetic learning paradigms. *Frontiers in human neuroscience*, 14:320, 2020.

[25] Y. Sasaki et al. The truth of the f-measure. 2007. *URL: https://www. cs. odu. edu/mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07. pdf [accessed 2021-05-26]*, 49, 2007.

[26] SteamDB. Most played VR Games Steam Charts.

[27] 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. *IEEE Access*, 11:9859–9875, 2023. doi: 10.1109/ACCESS.2023.3240071

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, p. 6000–6010. Curran Associates Inc.

[29] J. Wöbbeking. Beat Saber generated more revenue in 2021 than the next five biggest apps combined, Aug. 2022.

[30] Q.-S. Xu and Y.-Z. Liang. Monte carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56(1):1–11, 2001.

A Hyperparameters

- Input Shape: (1024×21)
- Embedding Size: 24
- Hidden Size: 128
- Number of Layers: 2
- Output Size: 1
- Learning Rate: 0.00002
- Epochs: 100
- Batch Size: 32

B Survey Questions & Response Distributions

For a full list of survey questions and the response distributions of each question, please see the preprint version of this paper: `https://arxiv.org/abs/2305.19198`.

C Response Correlations

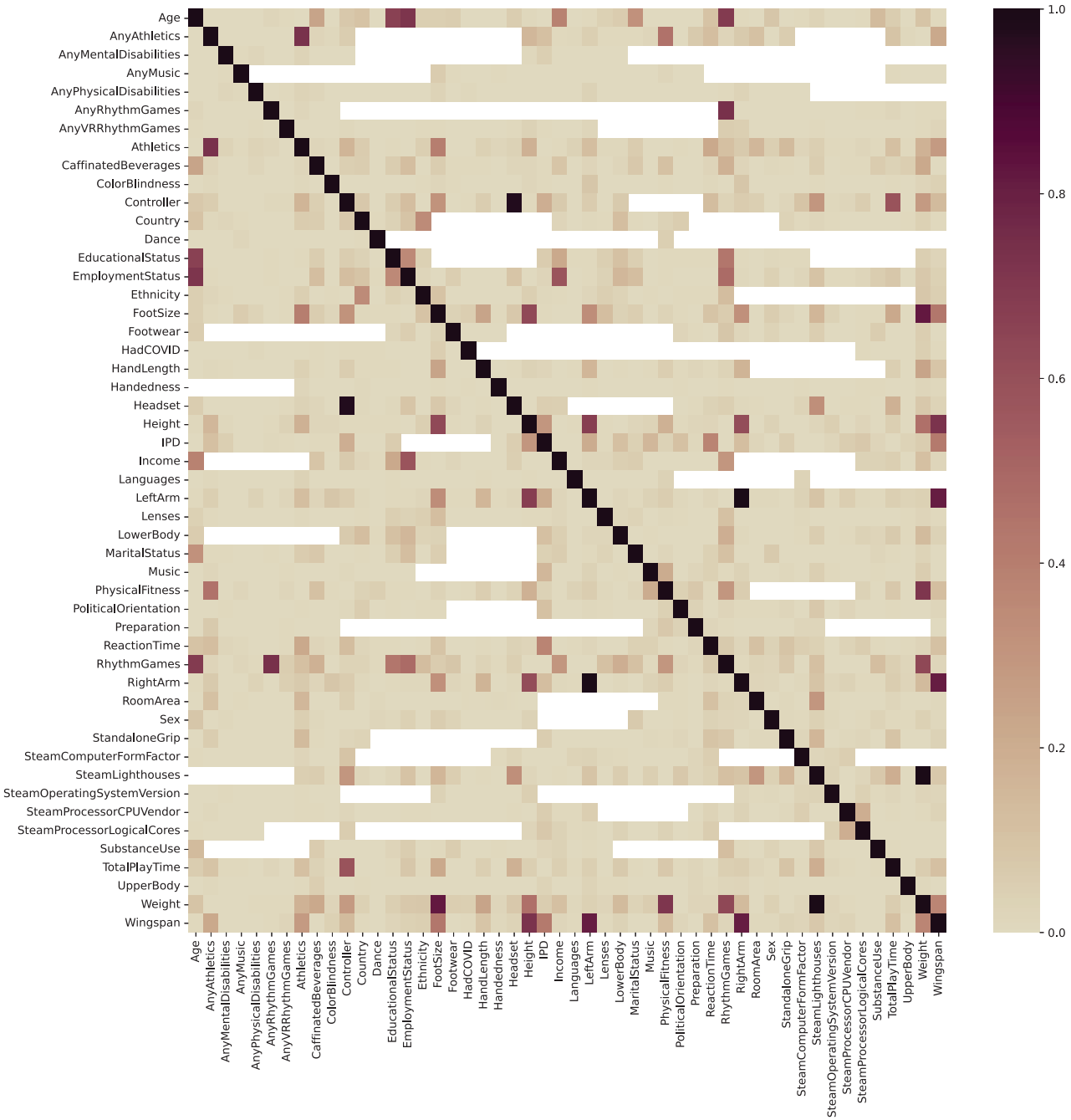

Figure 4: Correlation coefficient (R^2) between all pairs of attributes.