This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3372087

Berkeley Open Extended Reality Recordings 2023 (BOXRR-23): 4.7 Million Motion Capture Recordings from 105,000 XR Users

Vivek Nair, Wenbo Guo, Rui Wang, James F. O'Brien, Louis Rosenberg, and Dawn Song

Abstract—Extended reality (XR) devices such as the Meta Quest and Apple Vision Pro have seen a recent surge in attention, with motion tracking "telemetry" data lying at the core of nearly all XR and metaverse experiences. Researchers are just beginning to understand the implications of this data for security, privacy, usability, and more, but currently lack large-scale human motion datasets to study. The BOXRR-23 dataset contains 4,717,215 motion capture recordings, voluntarily submitted by 105,852 XR device users from over 50 countries. BOXRR-23 is over 200 times larger than the largest existing motion capture research dataset and uses a new, highly efficient and purpose-built XR Open Recording (XROR) file format.

Index Terms—Dataset, virtual reality, extended reality, motion capture, big data

1 Introduction

For decades, human motion capture (MoCap) recordings have been an important resource in a variety of fields, ranging from animation and computer-generated imagery (CGI) to authentication and human-computer interaction (HCI). Recently, the proliferation of extended reality (XR) devices has created a prominent new application for this data, with motion data being central to almost all XR and "metaverse" experiences. Since 2002, at least 25 motion capture datasets have been created based on laboratory studies of up to a few hundred users to facilitate research in this important domain.

An emerging area of interest for security and privacy researchers is the passive identification and authentication of XR users based on their movement patterns. Until recently, XR identification and authentication studies have been limited to a few hundred users due to the lack of large-scale human motion datasets. By contrast, studies involving traditional biometrics, such as fingerprints or facial recognition, often use datasets with 100,000 or more subjects [57].

In this paper, we introduce the BOXRR-23 dataset, which contains 4,717,215 motion capture recordings uploaded by 105,852 XR device users from over 50 countries. Our data is derived from two popular VR games, "Beat Saber" and "Tilt Brush." In addition to being more diverse and ecologically valid than laboratory studies, BOXRR-23 is over 200 times larger than the largest known public motion capture dataset (see §4). This dataset was recently used, for the first time, to demonstrate that XR motion data provides a biometric signal on par with fingerprints [34]. The identification result, published in *USENIX Security '23*, was made possible by this novel dataset. However, we envision the potential uses of this data may go far beyond security and privacy to include areas such as motion synthesis, human-computer interaction, and machine learning research.

In addition to assembling this dataset from three public sources and enriching it with additional metadata, we developed a new "Extended Reality Open Recording" (XROR) file format due to the lack of an existing standard format suitable for this use case. The XROR format is about 30% more space efficient than the original motion capture file formats, without loss of precision.

To help interested researchers evaluate this dataset, we provide documentation pursuant to a number of open standards, including Datasheets for Datasets [12] and Dataset Nutrition Labels [14]. Furthermore, we conducted a large-scale survey ($N = 1,006$) of the users contained in this dataset to better understand their demographics, the results of which are summarized herein.

2 Background

Since the 1990s, computerized motion tracking systems have been used for animation and CGI in a large number of popular movies, television series, and video games. A typical commercial motion capture solution uses optical tracking or inertial measurement units (IMUs) to measure the location of various body parts, with prices ranging from $10,000 to over $250,000 for a full-body tracking system. Conventional motion capture datasets have involved expensive laboratory studies with up to 300 subjects paid to perform a variety of tasks while wearing a professional motion capture setup.

Motion capture data is also central to the operation of extended reality (XR) systems, which include devices supporting augmented reality (AR), virtual reality (VR), and mixed reality (MR) technologies. XR has experienced a recent surge in attention and popularity with the release of self-contained VR devices like the Meta Quest and Apple Vision series. Most consumer-oriented virtual reality systems include a head-mounted display (HMD) and two hand-held controllers. The system uses either external or onboard sensors to measure the position and orientation of these devices in 3D space, providing six degrees of freedom (6DoF), captured at a rate of between 60 and 144 times per second. In essence, XR devices have recently become an affordable and widely-adopted form of motion tracking system.

The motion data generated by an XR device is used by a client-side application, such as "Beat Saber" or "Tilt Brush," to render auditory, visual, and haptic stimuli, creating an immersive 3D experience. In some cases, users capture and share recordings of this motion data to allow other users to "replay" the same virtual experience.

2.1 Beat Saber

"Beat Saber" [11], shown in Figure 1, is a VR rhythm game where players slice blocks representing musical beats with a pair of sabers they hold in each hand. It is the primary data source for the BOXRR-23 dataset. With over 6 million copies sold, Beat Saber is the most popular VR application of all time [58]. The game contains a number of "maps," which consist of an audio track and 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 saber, "bombs," which the player must avoid hitting with their sabers, and "walls," which the player must avoid with their head. The player is given a score based on their accuracy in completing these tasks. Reacting to these events typically requires users to deploy fast ballistic movements [6,56].

- *Vivek Nair is with UC Berkeley. E-mail: vcn@berkeley.edu*
- *Wenbo Guo is with Purdue University. E-mail: henrygwb@purdue.edu*
- *Rui Wang is with Carnegie Mellon. E-mail: ruiwang3@andrew.cmu.edu*
- *James F. O'Brien is with UC Berkeley. E-mail: job@berkeley.edu*
- *Louis Rosenberg is with Unanimous AI. E-mail: louis@unanimous.ai*
- *Dawn Song is with UC Berkeley. E-mail: dawnsong@berkeley.edu*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

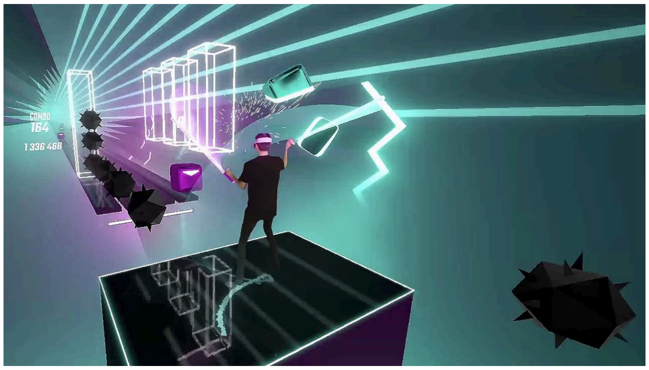

Fig. 1: “Beat Saber” – VR rhythm game.

While hundreds of maps are included in the base game, over 100,000 user-created maps can be played by installing open-source game modifications. Beat Saber enthusiasts may choose to install open-source leaderboard extensions in order to compete with other players to achieve a higher “rank” on the leaderboards for popular maps. Two of the most popular Beat Saber leaderboard services are “BeatLeader” [44] and “ScoreSaber” [48], with a combined 4 million scores being submitted to the platforms to date. When submitting a score to either of these services, users attach a motion capture recording of them playing the corresponding Beat Saber map, which is then made publicly available on the BeatLeader or ScoreSaber website to allow others to audit the legitimacy of the claimed score.

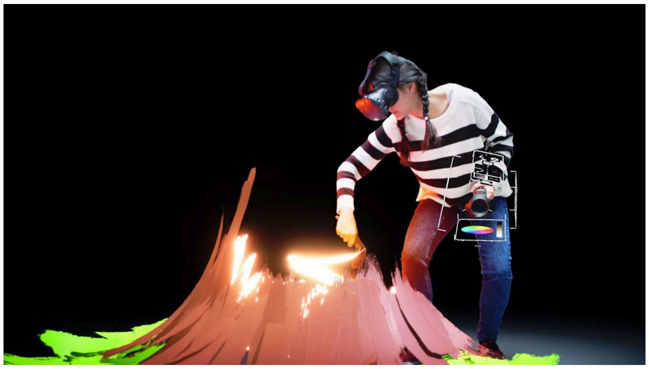

Fig. 2: “Tilt Brush” – VR painting app.

2.2 Tilt Brush

“Tilt Brush” [52], shown in Figure 2, is a VR painting game created by Google that allows users to create 3D virtual objects using a variety of brushes and tools. Users can then export their drawings in various file formats, along with a motion capture recording of them creating the object, allowing other users to re-watch the original painting process. From 2017 to 2021, Google hosted “Google Poly,” a free service for sharing virtual creations (and accompanying motion capture recordings) from Tilt Brush. After the shutdown of Google Poly in 2021, the “PolyGone” project [42] was created to host a free archive of over 50,000 user-submitted creations from Google Poly under a CC-BY license. Contrary to Beat Saber, Tilt Brush motion consists primarily of precise fine motor movements, providing a complementary data source.

3 Data Collection

Figure 3 shows the data collection process used to produce the BOXRR-23 dataset. We downloaded over 4.7 million publicly-available motion capture recordings stored on the BeatLeader, ScoreSaber, and PolyGone websites, and obtained additional metadata information, such as player experience levels and in-game events, from the public web APIs of Steam [51] and BeatSaver [45]. We then removed identifiable details like player IDs and pseudonyms to protect the identity of each user. Finally, we converted all recordings from their original formats into our purpose-built XROR format, described in §5. The sizes of each of the sources, and of the dataset, are summarized in Table 1. We performed this data collection process in April 2023 and have included all valid, non-corrupt recordings submitted to all three platforms between November 1st, 2017 and April 15th, 2023.

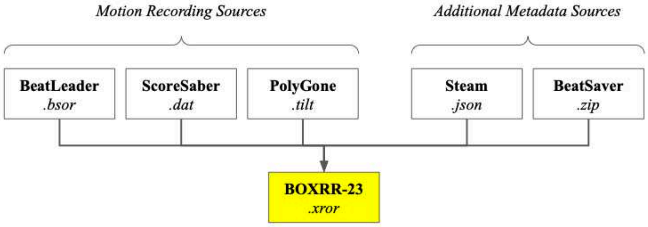

Fig. 3: Data collection/processing pipeline for BOXRR-23.

Table 1(A): Sources for data in BOXRR-23 dataset.

Source	Application	Users	Recordings	Format	Size
BeatLeader	Beat Saber	95,192	3,525,456	.bsor	6.25 TB
ScoreSaber	Beat Saber	55,331	1,136,581	.dat	1.44 TB
PolyGone	Tilt Brush	27,693	55,178	.tilt	1.87 TB

Table 1(B): Output characteristics of BOXRR-23 dataset.

Dataset	Users	Recordings	Format	Size
BOXRR-23 Dataset	105,852	4,717,215	.xror	4.71 TB

Table 1: BOXRR-23 characteristics and data sources.

4 Related Work

We searched for existing datasets relating to “motion capture,” “telemetry,” “VR motion,” “XR motion,” etc., on dataset hosting platforms like Kaggle, Zenodo, and Dryad, as well as for academic papers relating to motion capture data and experiments. The full set of search parameters used is included in the appendices. We found over 25 existing datasets containing human motion recordings, as summarized in Table 2.

Table 2(A): Current motion capture datasets outside XR.

Dataset	Organization	Year	Subjects	Recordings	Markers
BMLrub [54]	Ruhr Univ. Bochum	2002	111	3,061	41, 3DoF
HDM05 [32]	Max Planck Society	2007	4	215	41, 3DoF
CMU-MMAC [22]	Carnegie Mellon Univ.	2008	5	5	41, 3DoF
EYES Japan [30]	EYES Japan	2009	12	750	37, 3DoF
HumanEva [50]	Univ. of Toronto	2010	3	28	39, 3DoF
SFU MoCap [49]	Simon Fraser Univ.	2012	7	44	53, 3DoF
ACCAD [1]	Ohio State Univ.	2012	20	252	82, 3DoF
Sleight of Hand [15]	Trinity College Dublin	2012	1	62	91, 3DoF
Human3.6m [16]	Romanian Academy	2013	11	44	24, 3DoF
MoSh [24]	Max Planck Society	2014	19	77	87, 3DoF
MPI Limits [2]	Max Planck Society	2015	3	35	53, 3DoF
KIT MoCap [26]	Karlsruhe Inst. of Tech.	2016	232	2,925	50, 3DoF
Total Capture [55]	Univ. of Surrey	2017	5	37	53, 3DoF
AMASS [25]	Max Planck Society	2019	344	11,265	37, 3DoF
CMU MoCap [4]	Carnegie Mellon Univ.	2019	144	2,605	41, 3DoF
MoVi [13]	Queen’s Univ.	2021	90	1,890	12, 3DoF

Table 2(B): Current motion capture datasets inside XR.

Dataset	Organization	Year	Subjects	Recordings	Trackers
Behavioural Biometrics [40]	Bundeswehr Univ. Munich	2019	22	88	3, 6DoF
TTI [27]	Stanford Univ.	2020	511	511	3, 6DoF
Body Normalization [23]	Univ. of Duisburg-Essen	2021	16	48	3, 6DoF
Obfuscation [31]	Univ. of Central Florida	2021	60	120	3, 6DoF
Body Sway [5]	Purdue Univ.	2021	28	336	3, 6DoF
You Can’t Hide [53]	Univ. of Padova	2022	35	69	3, 6DoF
Motion Matching [43]	Univ. of Catalonia	2022	1	12	3, 6DoF
Personal Identifiability [28]	Stanford Univ.	2023	232	1856	3, 6DoF
Who is Alyx [47]	Univ. of Würzburg	2023	71	142	3, 6DoF

Table 2(C): Our new XR motion capture dataset.

Dataset	Organization	Year	Subjects	Recordings	Trackers
BOXRR-23	UC Berkeley, et al.	2023	105,852	4,717,215	3, 6DoF

Table 2: Comparison of BOXRR-23 with existing datasets.

The majority of these datasets come from conventional non-XR motion tracking systems, as listed in Table 2(A), while several originate from XR-based laboratory studies, listed in Table 2(B). The largest existing study contained 511 subjects [27], with a single session captured from each subject. By contrast, our dataset, summarized in Table 2(C), contains over 105,000 subjects and 4.7 million recordings from the three sources described in §3.

In addition to being over 200 times larger than the largest existing dataset, we found that all of the existing datasets come from a laboratory study in which participants used a small number of homogeneous devices and were generally physically present in a narrow geographical area. Thus, the BOXRR-23 dataset is more useful for obtaining a representative sample of XR users, as it originates from real XR users using their own devices in their own homes. As a result, it contains diverse data from over 40 types of XR devices, and includes users from over 50 countries around the world.

As evidenced by Table 2, BOXRR-23 is more comparable to existing XR datasets with a small number of 6DoF trackers than non-XR datasets with a large number of 3DoF markers. In applications where detailed full-body tracking is required, a conventional MoCap dataset may be more appropriate than an XR dataset like BOXRR-23.

5 XROR Format

As detailed in §3, the data included in the BOXRR-23 dataset was scraped from three separate public data sources (BeatLeader, ScoreSaber, and PolyGone), each using three separate custom file formats designed specifically for those platforms (.BSOR, .DAT, and .TILT, respectively, summarized in Table 3(A)). We felt that the experience of researchers consuming this dataset in the future would be improved if the recordings were all converted to a single file format that could be analyzed and ingested via a unified pipeline.

We began by evaluating open-source motion capture file formats such as .BVA, .BVH, and .MVNX. Unfortunately, we found that the existing formats were unsuitable for this database for a variety of reasons. Some formats, such as .BVA and .BVH, only have support fc motion data, and did not allow us to embed the rich metadata and even data streams we wished to include in the dataset. Others, like .MVN did support the inclusion of arbitrary metadata and event data stream but used an inefficient underlying text-based file format (.XML) tha would have caused the dataset to balloon to over 300 TB in size. Finall some proprietary formats did contain all of the necessary features i an efficient binary format, but were not open-source and required pai tools or licenses to utilize them. Overall, we found that none of th existing open-source file formats were unsuitable for this dataset.

A formal specification of the XROR format, using the BSON version of the JSON Schema notation, is here: https://rdi.berkeley.edu/metaverse/boxrr-23/dict.json.

Table 3(A): Source file formats for motion data.

Format	Metadata	Motion Data	Event Data	Compression	Avg. Size
.tilt	✓	✓	✓		33.89 MB
.bsor	✓	✓	✓		1.77 MB
.dat	✓	✓	✓		1.27 MB

Table 3(B): Existing general file formats for motion data.

Format	Metadata	Motion Data	Event Data	Compression	Avg. Size
.mvnx	✓	✓	✓		61.90 MB
.bvh		✓			25.79 MB
.bva		✓			13.98 MB

Table 3(C): Proposed new open file format for motion data.

Format	Metadata	Motion Data	Event Data	Compression	Avg. Size
.xror	✓	✓	✓	✓	0.99 MB

Table 3: Comparison of XROR with existing formats.

To address the issues with existing open-source file formats, we introduce the new "Extened Reality Open Recording (XROR)" file format. XROR files contain metadata as well as rich event and motion data streams, and are based internally on BSON (Binary JSON), a flexible, widely-supported format with libraries in dozens of languages. Metadata is stored as JSON key-value pairs, while event data and motion data streams are converted to 2D floating-point arrays and compressed using fpzip, a lossless compressor of multidimensional floating-point arrays designed by Lawrence Livermore National Laboratory specifically for the efficient storage and transmission of scientific datasets.

To evaluate the relative efficiency of our new format, we converted a portion of our dataset into a variety of existing open formats, summarized in Table 3(B), as well as our proposed XROR format, as shown in Table 3(C). Even compared to the original, purpose-built formats shown in Table 3(A), XROR achieves space savings of at least 30% with no loss in precision due to the use of fpzip compression.

Due to the advantages of our new XROR format over the existing alternatives, the entire BOXRR-23 dataset is offered exclusively as XROR files. To help researchers process this format, we have provided open-source tools to parse XROR files, and convert them to and from a variety of formats (e.g., TILT, BSOR, DAT, and JSON).

6 Recording Contents

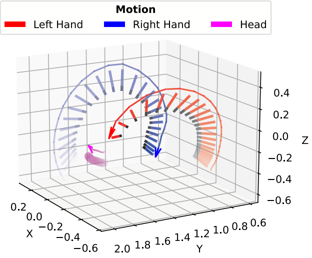

Fig. 4: "Beat Saber" motion.

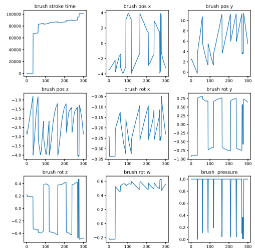

Fig. 5: "Tilt Brush" motion.

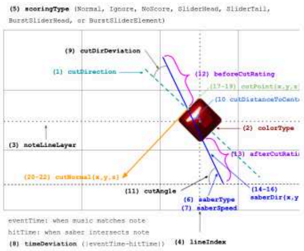

Fig. 6: "Beat Saber" event.

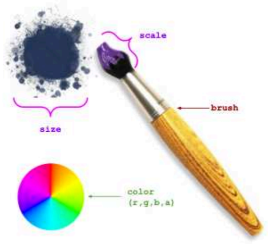

Fig. 7: "Tilt Brush" event.

Figures 4–7 illustrate the typical contents of each recording in the BOXRR-23 dataset. The following data is included in each recording:

1. **Metadata**. A variety of metadata is included with each entry, including anonymized user IDs, hardware and software descriptions, and virtual environment and activity descriptions.
2. **Motion data**. Recordings principally consist of motion data captured in 6DoF at between 60 Hz and 144 Hz. Beat Saber recordings include head and hand motion data (see Fig. 4), while Tilt Brush recordings include brush motion and pressure data (see Fig. 5).
3. **Event data**. Motion data is accompanied by rich contextual information about events occurring in the virtual world. This includes information about the in-game objects and obstacles in the case of Beat Saber (see Fig. 6), and about each brush stroke in the case of Tilt Brush (see Fig. 7).

Detailed data examples of Beat Saber and Tilt Brush recordings in the BOXRR-23 dataset are provided in the supplemental materials.

7 Access Instructions

Researchers interested in using the BOXRR-23 dataset are invited to visit https://rdi.berkeley.edu/metaverse/boxrr-23/. The permanent DOI is https://doi.org/10.25350/B5NP4V. For ease of access, the dataset has been split into 106 .zip files, each containing up to 1,000 users. Each user is represented by a folder containing .xror recordings from that user.

We developed the licensing terms for this dataset in conjunction with the Committee for Protection of Human Subjects (CPHS) and Intellectual Property & Industry Research Alliances (IPIRA) groups at UC Berkeley, with the chief goal of protecting the human subjects contained in this dataset. The dataset is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license, and is additionally subject to a data use agreement (DUA) that prohibits unethical uses of the data, such as attempts to deanonymize the subjects. Online access to the dataset is automatically granted upon agreeing to the license and DUA.

8 Intended Use Cases

As discussed above, known uses of this dataset are primarily in the authentication and biometrics domain. However, there are a number of interesting envisioned uses for this dataset within the VR community, beyond security and privacy research.

8.1 Notable Known Uses

Until recently, this dataset has only been available for internal use at UC Berkeley. Thus far, we have published four papers using this dataset in the XR security and privacy domain:

- We conducted a study that uniquely identified over 55,000 VR users based on their head and hand motion [34]. By using the BOXRR-23 dataset, this study was over 200 times larger than the next largest VR identification study, and the first to demonstrate parity with biometrics like fingerprints.
  - Result: After training a classification model on 5 minutes of data per person, a user can be uniquely identified amongst the entire pool with 94.33% accuracy from 100 seconds of motion.
  - Availability: The source code and documentation required to replicate this result using the BOXRR-23 dataset can be found at https://github.com/metaguard/identification.
- In another study, we combined the BOXRR-23 dataset with a survey to demonstrate that a large number of sensitive data attributes can be inferred from VR users based on motion alone [37].
  - Result: Using simple machine learning models, over 35 private data attributes could accurately and consistently be inferred from VR users using head and hand motion data alone.
  - Availability: The source code and documentation required to replicate this result using the BOXRR-23 dataset can be found at https://github.com/metaguard/profiling.
- In a third paper, we presented "MetaGuard," [33] a differential privacy-based tool for protecting user data privacy in the metaverse, which we evaluated using the BOXRR-23 dataset.
  - Result: We show a significant degradation of attacker capabilities when using MetaGuard.
  - Availability: The source code and documentation required to replicate this result using the BOXRR-23 dataset can be found at https://github.com/metaguard/metaguard.
- In a fourth study, we presented "Deep Motion Masking," [36] a machine learning architecture for anonymizing VR motion data, which we trained and evaluated using the BOXRR-23 dataset.
  - Result: Through a large-scale user study (N=182), we demonstrate that our method is significantly more usable and private than existing VR anonymity systems.
  - Availability: The source code and documentation required to replicate this result using the BOXRR-23 dataset can be found at https://github.com/metaguard/metaguardplus.

8.2 Future Directions

While the dataset has primarily been used in the security and privacy domain, we can envision a number of additional interesting applications for this data. Historically, motion capture data has primarily been used for computer graphics, animation, and CGI, and our data could also be used in this domain. For example, it could be used to train large-scale generative machine learning models for natural human motion synthesis. It may also be of interest to researchers studying human-computer interaction in XR (e.g., researchers could use the data to investigate interaction patterns likely to cause discomfort or injury).

One area of active research that is relevant to our dataset is the inference of full-body pose information from sparse tracking inputs. Researchers have demonstrated the ability to recover full-body motion data from the motion of a few tracked points [7, 17]. Using these techniques, the sparse tracking data offered by our dataset could be used to recover inferred full-body motion for various uses.

Furthermore, the dataset contains numerous labels, including anonymized user IDs, hardware and software descriptions, and virtual environment and activity descriptions, that can be used to construct novel classification and regression tasks. For example, a very interesting use of the Tilt Brush portion of the dataset could be to use the brushstroke motion data to infer the title or description of the drawing, which are provided in the metadata as potential labels.

Finally, this dataset presents a challenging and unique opportunity for theoretical machine learning research, because it consists of long, sequential data, with sequence lengths often in excess of 100,000. Most existing deep learning algorithms are not well equipped to handle sequential data of this size. Currently, our dataset is a rare instance of a task in which classical ML algorithms seem to outperform deep learning methods [34]. Developing models that can accurately and efficiently ingest the data contained in this dataset may require theoretical advances in machine learning techniques.

9 Preliminary Analysis

In this section, we offer a preliminary analysis of the BOXRR-23 dataset, in which we primarily summarize the metadata associated with each recording to complement the existing uses of the dataset (§8.1) and benchmarking results (§11). The dataset contains 4,717,215 recordings from 105,852 users. Recordings vary in length from one second to over an hour, with an average length of about three minutes. Over 97% of recordings in the dataset are between one and seven minutes in length.

The recordings are also not evenly distributed across the 105,852 users: the most prolific individual users have over 1,000 recordings each in the dataset, while 90% of users have 15 or fewer recordings. The top 1,000 users alone account for nearly 500,000 recordings or about 10% of the dataset, despite being less than 1% of the users.

Additionally, over 30 different models of extended reality devices are present in the dataset. However, as illustrated in Fig. 8, Oculus Quest 2 devices are by far the most popular, with all Oculus models together representing about 66% of the dataset. Our metadata also includes the runtime environment of each recording, with about 75% of recordings using SteamVR (including Quest Link) and about 25% using the Oculus runtime. Neither the Oculus Quest 3 nor the Apple Vision Pro were broadly available at the time this dataset was collected.

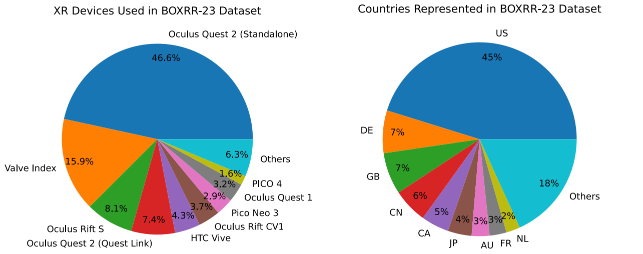

Fig. 8: Distribution of devices and countries in BOXRR-23 recordings.

Furthermore, IP-based geolocation estimates are available for about 70% of the users in the dataset. While over 50 different countries are represented, users from the United States account for nearly half of the dataset as shown in Fig. 8. About 66% of the dataset originates from the top five countries alone, as discussed further in §12.

Finally, basic anthropometric data can be observed from the dataset. For example, 4.3% of users have configured their device in left-handed mode, while 95.7% use the default right-handed mode. Globally, it is believed that approximately 10% of people are left-handed, indicating that left-handed users are either underrepresented in our dataset, or largely choose to leave their device in right-handed mode. Furthermore, the median height setting of users in the dataset is approximately 1.7 meters, which closely aligns with the true global median height.

10 Population Survey

To shed additional light on the demographics of the users within our dataset, we conducted a large-scale online survey of VR users. The survey contained about 50 questions and received 1,006 responses, of which 830 users were present in the BOXRR-23 dataset. It was conducted in coordination with BeatLeader and other Beat Saber organizations, and thus did not reach the 1% of BOXRR-23 users from Tilt Brush. The full results of this survey are available online [38], with primary demographics summarized in Figure 9 below.

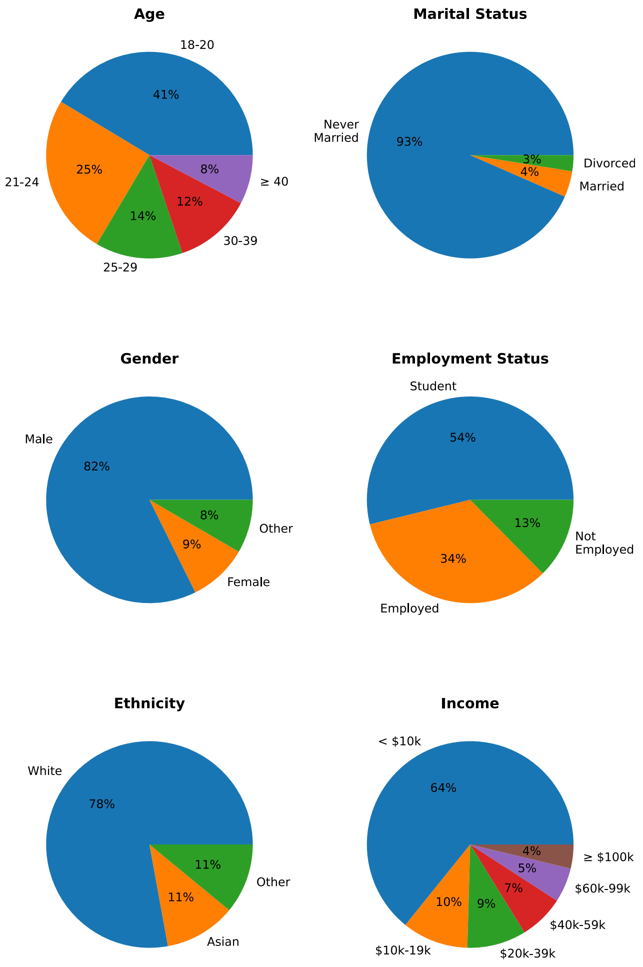

Fig. 9: Survey results from 830 users in the BOXRR-23 dataset.

11 Benchmark Result

In this section, we present a baseline motion-based XR user identification result to demonstrate the potential of the BOXRR-23 dataset for large-scale user identification tasks, above and beyond the existing deployments detailed in §8.1. We describe the basic principles behind existing VR identification models and then show that with the large volume of data available in BOXRR-23, models can now be trained that are far more robust and capable than those discussed in prior work.

11.1 Prevailing Architectures

At present, most existing papers on VR user identification utilize classical machine learning models, such as those based on the Random Forest [3] and LightGBM [18] architectures. The motivation for using these models over more powerful deep learning approaches is that deep learning typically requires a significantly larger volume of data to successfully train and converge, whereas tree-based architectures can produce generalizable classifiers with fewer samples per user.

On the other hand, the sequential time-series format of VR motion data streams is not a natural fit for tree-based models, which require a one-dimensional tabular data format. As such, prior works suggest deliberate feature engineering to convert motion data streams into tabular samples by using summary statistics to eliminate the time dimension.

Specifically, Pfeuffer et al. [41] suggest dividing motion data into one-second chunks, and then converting each chunk into a flat feature vector by taking four statistics (min, max, mean, and standard deviation) across each tracked dimension. Miller et al. [29] use a very similar approach, but also include the median of each axis. Moore et al. [31] use identical features to Miller, while Nair et al. [35] use similar features but add contextual data specific to the VR application. At a high level, many prior works have found the basic idea of summarizing one-second chunks of motion to be highly effective. Still, these approaches remain a concession forced by not having enough data to use deep learning. Now, having access to the massive BOXRR-23 dataset, we are motivated to produce similar identification experiments using deep learning architectures in order to achieve better identification performance.

11.2 User Identification Benchmark

We now demonstrate how an LSTM architecture can be used to drive improvements in motion-based identification accuracy, provided a large amount of training data per user is available. Using the BOXRR-23 dataset, we first found the 500 users for which the greatest number of individual recordings were available. For these top 500 users, an average of 821 recordings were available per user, with each recording averaging about three minutes in length. We used the 500 most recent recordings of each user for our evaluation, with 400 of these being used for training, 50 for validation, and the remaining 50 being used for testing. Only the first 30 seconds of each recording were utilized, and recordings were normalized to a constant 30 frames per second by using a numerical linear interpolation for positional coordinates and a spherical linear interpolation for orientation quaternions.

To evaluate the performance of the LSTM funnel architecture on this particular dataset, we implemented a two-layer LSTM architecture in Keras v2.10.1 [19] and trained it for 500 epochs on the described dataset using the Adam optimizer [20] with a learning rate of 0.001. The validation dataset was used for early stopping after 25 epochs of no improvement. For the sake of comparison, we also trained and tested several previously proposed identification model architectures using the same dataset, the results of which were as follows:

- Our new LSTM funnel architecture achieves a per-sample accuracy of 98.12% and a per-user accuracy of 100.00%.
- The Nair et al. [35] architecture achieves a per-sample accuracy of 71.66% and a per-user accuracy of 100.00%.
- The Miller et al. [28] architecture achieves a per-sample accuracy of 56.59% and a per-user accuracy of 97.60%.

As evidenced by the above results, our architecture substantially exceeds the identification performance of the most notable prior models when using identical datasets. This, on its own, is not entirely surprising, given that we used over three hours of training data per user to perform this demonstration, which also exceeds all prior works; the previously proposed models and featurization approaches were not designed to take full advantage of this volume of data. However, due to the volume of training data used, the model achieves an unprecedented level of robustness to reductions in input dimensionality:

- The original representation with the full 21 features ($\{head, left\_hand, right\_hand\} \times \{x,y,z,i,j,k,w\}$) gives a sample accuracy of 98.12% and a user accuracy of 100.00%.
- Removing the head, the remaining 14 features ($\{left\_hand, right\_hand\} \times \{x,y,z,i,j,k,w\}$) reduce sample accuracy to 94.76% (and still 100% user accuracy).
- Using only hand rotations, the remaining 8 features ($\{left\_hand, right\_hand\} \times \{i,j,k,w\}$) give a sample accuracy of 93.42% and a user accuracy of 100.00%.
- Using only left hand rotations, the remaining 4 features ($\{left\_hand\} \times \{i,j,k,w\}$) still result in a sample accuracy of 92.77% and a user accuracy of 100.00%.
- Using only left hand rotational magnitude, the single feature ($\{left\_hand\} \times \{w\}$) still results in a sample accuracy of 84.23% and a user accuracy of 100.00%.

In other words, by observing just the magnitude of the rotation of one hand of a user for a period of 30 seconds, the model can still correctly identify the user out of 500 options with nearly 85% accuracy, provided it was first trained on over three hours of data for each user.

Previously, obtaining over three hours of motion capture data from each of 500 users would have been infeasible in the context of a laboratory study, with prior studies either having 500 users but less than ten minutes of data per user [27], or having hours of data per user but less than 100 users [47]. As the above results demonstrate, having both a large number of users and a large amount of data per user is critical to enabling highly robust motion-based identification in VR.

If extended reality truly replaces existing mobile devices as a default method of human-computer interaction for millions of users in the near future, having multiple hours of cumulative time spent using XR devices may soon come to represent an average or even below-average usage pattern. The BOXRR-23 dataset provides the first major opportunity for researchers to understand the potential implications of this large-scale motion data were it to become broadly available.

12 Limitations

As may be evident by the survey results provided in §10, the users included in our dataset are not necessarily representative of a general population. For example, the dataset consists primarily of white and male subjects. While the subjects are demographically similar to the overall population of VR device users [8], they consist entirely of users who chose to upload a BeatSaber performance or TiltBrush drawing to a public platform. As such, we believe enthusiast or expert-level users are likely to be overrepresented in the dataset. However, for the same reason, the dataset likely contains far more geographic diversity than existing laboratory-based datasets. Furthermore, the data is derived from just two VR applications, Beat Saber and Tilt Brush, with almost 75% of the users and 99% of the recordings being from Beat Saber alone. Overall, researchers should be cautious when attempting to use this dataset to draw conclusions about larger populations than the ones directly included. When attempting to use BOXRR-23 to draw conclusions about broader populations, researchers should follow best practices for accounting for sampling bias in datasets [21, 39].

Additionally, there are some risks associated with the dataset being derived from ordinary XR users. Some metadata values, such as Beat Saber song titles or Tilt Brush drawing descriptions, may contain objectionable content due to their user-submitted nature. Metadata constituting user-configured settings like height and handedness should be considered self-reported, and are subject to the typical response biases associated with self-reported values. Finally, because the data is from "the wild" rather than a laboratory study, it originates from a wide variety of heterogeneous XR devices and physical environments, and may include more noise and tracking errors than a lab-created dataset.

13 Ethical Considerations

Because our dataset consists entirely of motion capture recordings from human subjects, significant attention was given to ethics throughout the process of designing and collecting the dataset. Our collection of this dataset was approved and overseen by the UC Berkeley Office for Protection of Human Subjects (OPHS), an OHRP-certified Institutional Review Board (IRB), approved as protocol #2023-03-16120. .

We note that in producing this dataset, the authors had no direct contact with human subjects. Instead, our data is derived from three public sources. All data utilized in this study was already broadly, publicly available, to any person in the world with an internet connection, without the need for permissions, credentials, authentication, or any special tools or applications, via the websites of ScoreSaber, BeatLeader, and PolyGone. No new data is being made accessible to the public in the publication of this dataset; our contribution is in finding, scraping, aggregating, reprocessing, enriching, and distributing this existing data, and in surveying the underlying population.

Despite the public nature of the data and the IRB approval, we chose to obtain written permission from ScoreSaber, BeatLeader, and PolyGone before proceeding out of an abundance of caution and respect for the communities from which this data originates. We did not begin collecting data until authorized to do so by these communities, and sought their input throughout the collection process.

Users of the ScoreSaber, BeatLeader, and PolyGone platforms must voluntary install custom software to share their motion recording data with these platforms. They are fully aware of the nature of the data being shared, as uploading and publicly sharing XR data is the explicit purpose of these platforms. They also consent to their recordings being made publicly available in the privacy policies of these platforms. For example, the BeatLeader Privacy Policy, which can be found at `https://www.beatleader.xyz/privacy`, states that "Replays may contain personally identifiable information... Your data, including associated personally identifiable information, will be broadly publicly available to anyone with an internet connection via the BeatLeader website." Users of Google Poly (and PolyGone) consent to making their data publicly available under a CC-BY license.

Beyond consenting to the publication of their data in privacy policies and license agreements, we made further attempts to notify users of their involvement in academic research. Because users authenticate with these platforms via OAuth, their contact information is not known to the platforms, making direct consultation infeasible. However, we worked in collaboration with the BeatLeader team to inform users of their inclusion in academic research via their website and the official social media channels of the platform, and to develop an opt-out mechanism.

Although users knowingly consented to the public availability of their motion data, we took additional steps to protect the privacy of data subjects. First, all known explicit identifiers, such as usernames and user IDs, have been removed from the dataset. No potentially sensitive information, such as protected health information, is included in the data or metadata. Second, the dataset is offered under a data use agreement (DUA) that prohibits researchers from attempting to deanonymize the users, or to infer private attributes of the users that may be deemed sensitive. We followed the strictest PII data handling guidelines offered by our institution throughout the dataset collection process to preclude the accidental release of non-anonymized data.

Participants originally submitted their motion data to the ScoreSaber, BeatLeader, and PolyGone platforms for purposes other than academic research. Namely, they chose to make their data freely publicly available for reasons such as competitive e-sports or collaborative artwork; as such, users were not compensated for their original submissions, nor for their inclusion in the dataset. Moreover, any participant risks associated with the use of an extended reality device would have been realized by the users regardless of the later inclusion of the resultant motion recordings in this dataset. The scraping and redistribution of publicly-available online data is a highly common and widely accepted practice within the machine learning community [9, 46].

While it is impossible to entirely eliminate the risks associated with a new dataset, we believe the additional risk posed by our dataset is minimal in light of the fact that all of the included data was already public.

On the other hand, the data has the potential to facilitate significant advances in fields like graphics, HCI, XR, AI/ML, and computer security and privacy. We have taken significant steps to mitigate the potential harms of this dataset while maximizing its utility for beneficial research. Overall, we believe this research constitutes a net benefit to the subjects whose data was included by shedding light on the implications of the motion capture data which they have already, independently chosen to publish. For instance, security and privacy research using this dataset benefits society by highlighting the magnitude of the VR privacy threat and motivating future work on countermeasures.

14 Conclusion

We have presented the BOXRR-23 dataset, a 4.7 TB dataset of extended reality motion capture recordings from users around the world. Unlike existing motion capture datasets, BOXRR-23 is derived from recordings submitted by participants using their own XR devices, rather than a laboratory setup. As a result, it contains over 200 times more users, and over 400 times more recordings, than all known comparable datasets, while simultaneously being more diverse and ecologically valid.

The two XR applications included in BOXRR-23, Beat Saber and Tilt Brush, provide highly complementary motion data. Beat Saber consists almost entirely of fast ballistic movements while Tilt Brush consists almost entirely of fine motor movements, each controlled by a separate part of the brain [10]. By combining these sources, BOXRR-23 provides researchers a diverse collection of motion patterns.

For the first time, BOXRR-23 allows the identifiability of human motion data to be directly compared with biometrics like fingerprints and facial recognition, which have long enjoyed large public datasets. Our benchmarking results show that the massive scale of the BOXRR-23 dataset enables the use of deep learning for XR identification tasks, providing significant increases in identification accuracy and robustness. As such, we hope to see new advances in passive authentication mechanisms and privacy-preserving systems for XR, in addition to potential deployments in fields ranging from graphics and animation to usability and human-computer interaction.

In addition to identifying three new sources of motion data not previously widely known to academic researchers, we contributed a new XROR format to enable the efficient storage and transmission of this data. XROR is approximately 30% more efficient than the three original data formats, without any loss in precision, while also being more versatile than most existing open-source formats. Documentation for our dataset is offered according to widely-recognized open standards, including Datasheets for Datasets [12] and Dataset Nutrition Labels [14]. We also conducted a large survey of over 800 users present in the dataset to help researchers understand its demographic constituency.

As advances in extended reality allow this technology to reach increasingly large audiences, human motion data will remain vital to the operation XR and "metaverse" systems for the foreseeable future. In particular, augmented reality (AR) technology promises to be the next major medium of human-computer interactions, potentially even replacing the use of mobile devices such as smartphones. If this reality comes to pass, it is vital that we improve our understanding of the uses and implications of the motion data that these devices are designed to generate. We look forward to seeing future work that deploys the BOXRR-23 dataset to advance public knowledge in a variety of important fields, and to drive improvements to XR and metaverse experiences that benefit the field of extended reality as a whole.

Acknowledgments

This work was supported in part by the Minderoo foundation, the National Science Foundation (NSF), the National Physical Science Consortium (NPSC), the Fannie and John Hertz Foundation, and the Berkeley Center for Responsible, Decentralized Intelligence (RDI).

Availability

The identification benchmark code is available here: `https://github.com/MetaGuard/MetaGuardPlus/tree/main/motivation`

References

[1] Accad mocap system and data. 2

[2] I. Akhter and M. J. Black. Pose-conditioned joint angle limits for 3d human pose reconstruction. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1446–1455, 2015. doi: 10.1109/CVPR.2015.7298751 2

[3] L. Breiman. Random forests. *Machine learning*, 45:5–32, 2001. doi: 10.1023/A:1010933404324 5

[4] Cmu graphics lab motion capture database. 2

[5] S. Dent, K. Burger, S. Stevens, B. Smith, and J. Streepey. The effect of music on body sway when standing in a moving virtual environment. *PLOS ONE*, 16:e0258000, 09 2021. doi: 10.1371/journal.pone.0258000 2

[6] S. Douglas and A. Mithal. *The Ergonomics of Computer Pointing Devices*. Applied Computing. Springer London, 2012. 1

[7] Y. Du, R. Kips, A. Pumarola, S. Starke, A. Thabet, and A. Sanakoyeu. Avatars grow legs: Generating smooth human motion from sparse tracking inputs with diffusion model, 2023. 4

[8] J. Durbin. Report: Vive Users Are 95 Percent Male And Spend 5 Hours Per Week in VR, Feb. 2017. 6

[9] L. Fan, G. Wang, Y. Jiang, A. Mandlekar, Y. Yang, H. Zhu, A. Tang, D.-A. Huang, Y. Zhu, and A. Anandkumar. Minedojo: Building open-ended embodied agents with internet-scale knowledge. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. 6

[10] C. Fromm and E. V. Evarts. Relation of motor cortex neurons to precisely controlled and ballistic movements. *Neuroscience letters*, 5(5):259–265, 1977. 7

[11] B. Games. Beat Saber. `https://beatsaber.com/`. 1

[12] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumeé III, and K. Crawford. Datasheets for Datasets. *arXiv:1803.09010 [cs]*, Jan. 2020. 1, 7

[13] S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. J. Cook, G. Blohm, and N. F. Troje. MoVi: A large multi-purpose human motion and video dataset. *PLOS ONE*, 16(6):e0253157, jun 2021. doi: 10.1371/journal.pone.0253157 2

[14] S. Holland, A. Hosny, S. Newman, J. Joseph, and K. Chmielinski. The dataset nutrition label: A framework to drive higher data quality standards, 2018. 1, 7

[15] L. Hoyet, K. Ryall, R. McDonnell, and C. O'Sullivan. Sleight of hand: Perception of finger motion from reduced marker sets. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '12, p. 79–86. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2159616.2159630 2

[16] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, 2014. doi: 10.1109/TPAMI.2013.248 2

[17] J. Jiang, P. Streli, H. Qiu, A. Fender, L. Laich, P. Snape, and C. Holz. Avatarposer: Articulated full-body pose tracking from sparse motion sensing, 2022. 4

[18] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017. 5

[19] Keras: Deep learning for humans. 5

[20] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[21] B. Koch, E. Denton, A. Hanna, and J. G. Foster. Reduced, reused and recycled: The life of a dataset in machine learning research, 2021. 6

[22] F. D. la Torre, J. K. Hodgins, A. W. Bargteil, X. Martin, J. R. Macey, A. T. Collado, and P. Beltran. Guide to the carnegie mellon university multimodal activity (cmu-mmac) database, 2008. 2

[23] 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 2

[24] M. Loper, N. Mahmood, and M. J. Black. Mosh: Motion and shape capture from sparse markers. *ACM Trans. Graph.*, 33(6), nov 2014. doi: 10.1145/2661229.2661273 2

[25] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pp. 5442–5451, Oct. 2019. 2

[26] C. Mandery, O. Terlemez, M. Do, N. Vahrenkamp, and T. Asfour. The kit whole-body human motion database. In *2015 International Conference on Advanced Robotics (ICAR)*, pp. 329–336, 2015. doi: 10.1109/ICAR.2015.7251476 2

[27] M. Miller, F. Herrera, H. Jun, J. Landay, and J. Bailenson. Personal identifiability of user tracking data during observation of 360-degree vr video. *Scientific Reports*, 10, 10 2020. doi: 10.1038/s41598-020-74486-y 2, 3, 6

[28] M. R. Miller, E. Han, C. DeVeaux, E. Jones, R. Chen, and J. N. Bailenson. A large-scale study of personal identifiability of virtual reality motion over time, 2023. 2, 5

[29] R. Miller, N. Banerjee, and S. Banerjee. Within-system and cross-system behavior-based biometric authentication in virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 311–316, 03 2020. doi: 10.1109/VRW50115.2020.00070 5

[30] mocapdata.com. 2

[31] 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 2, 5

[32] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber. Documentation mocap database hdm05, 06 2007. 2

[33] V. Nair, G. M. Garrido, and D. Song. Going incognito in the metaverse, 2023. 4

[34] 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. 1, 4

[35] 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. In *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 895–910. USENIX Association, Anaheim, CA, Aug. 2023. 5

[36] V. Nair, W. Guo, J. F. O'Brien, L. Rosenberg, and D. Song. Deep motion masking for secure, usable, and scalable real-time anonymization of virtual reality motion data, 2023. 4

[37] V. Nair, C. Rack, W. Guo, R. Wang, S. Li, B. Huang, A. Cull, J. F. O'Brien, L. Rosenberg, and D. Song. Inferring private personal attributes of virtual reality users from head and hand motion data, 2023. 4

[38] V. Nair, V. Radulov, and J. F. O'Brien. Results of the 2023 census of beat saber users: Virtual reality gaming population insights and factors affecting virtual reality e-sports performance, 2023. 5

[39] T. P. Pagano, R. B. Loureiro, F. V. N. Lisboa, G. O. R. Cruz, R. M. Peixoto, G. A. de Sousa Guimarães, L. L. dos Santos, M. M. Araujo, M. Cruz, E. L. S. de Oliveira, I. Winkler, and E. G. S. Nascimento. Bias and unfairness in machine learning models: a systematic literature review, 2022. 6

[40] 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 2

[41] 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 5

[42] Polygone Art. 2

[43] J. L. Ponton, H. Yun, C. Andujar, and N. Pelechano. Combining Motion Matching and Orientation Prediction to Animate Avatars for Consumer-Grade VR Devices. *Computer Graphics Forum*, 41(8):107–118, 2022. doi: 10.1111/cgf.14628 2

[44] V. Radulov. BeatLeader. 2

[45] V. Radulov. BeatSaver. 2

[46] M. M. Rahman, D. Balakrishnan, D. Murthy, M. Kutlu, and M. Lease. An information retrieval approach to building datasets for hate speech detection. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. 6

[47] C. Schell, F. Sieper, L. Schach, and M. E. Latoschik. cschell/who-is-alyx: v2.0, Feb. 2023. doi: 10.5281/zenodo.7663984 2, 6

[48] ScoreSaber. 2

[49] Sfu motion capture database. 2

[50] L. Sigal, A. Balan, and M. Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, 87:4–27, 03 2010. doi: 10.1007/s11263-009-0273-6 2

[51] Steam. 2

[52] Tilt Brush by Google. 2

[53] 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, 2022. 2

[54] N. F. Troje. Decomposing biological motion: a framework for analysis and synthesis of human gait patterns. *Journal of vision*, 2 5:371–87, 2002. 2

[55] M. Trumble, A. Gilbert, C. Malleson, A. Hilton, and J. Collomosse. Total capture: 3d human pose estimation fusing video and inertial sensors. In *2017 British Machine Vision Conference (BMVC)*, 2017. 2

[56] S. N. P. Vitaladevuni. *Human Movement Analysis: Ballistic Dynamics, and Edge Continuity for Pose Estimation*. PhD thesis, University of Maryland at College Park, USA, 2007. AAI3297224. 1

[57] C. L. Wilson. Biometric accuracy standards. 1

[58] J. Wöbbeking. Beat Saber generated more revenue in 2021 than the next five biggest apps combined, Aug. 2022. 1