Identifying Users by Their Hand Tracking Data in Augmented and Virtual Reality

Jonathan Liebers[a,*], Sascha Brockel[a], Uwe Gruenefeld[a] and Stefan Schneegass[a]

[a]University of Duisburg-Essen, Schuetzenbahn 70, 45127 Essen, Germany

ARTICLE HISTORY
Compiled October 20, 2022

ABSTRACT
Nowadays, Augmented and Virtual Reality devices are widely available and are often shared among users due to their high cost. Thus, distinguishing users to offer personalized experiences is essential. However, currently used explicit user authentication (e.g., entering a password) is tedious and vulnerable to attack. Therefore, this work investigates the feasibility of implicitly identifying users by their hand tracking data. In particular, we identify users by their uni- and bimanual finger behavior gathered from their interaction with eight different universal interface elements, such as buttons and sliders. In two sessions, we recorded the tracking data of 16 participants while they interacted with various interface elements in Augmented and Virtual Reality. We found that user identification is possible with up to 95 % accuracy across sessions using an explainable machine learning approach. We conclude our work by discussing differences between interface elements, and feature importance to provide implications for behavioral biometric systems.

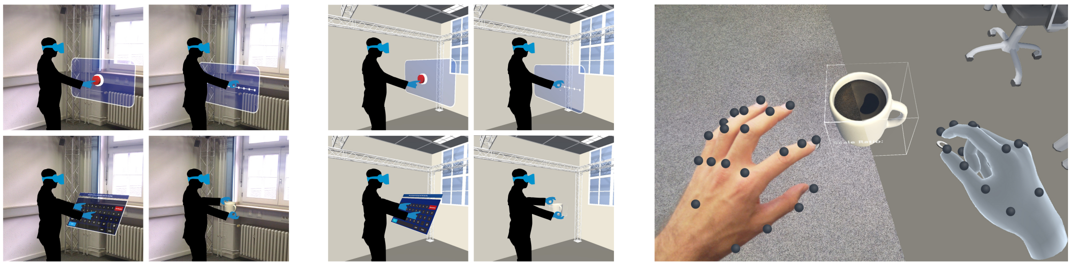
A **video abstract** of this work is available online at: `https://identifying-users-by-hand-tracking-data.hcigroup.de`

KEYWORDS
Augmented Reality; Behavioral Biometrics; Hand Tracking; User Identification; Virtual Reality

1. Introduction

Finger and hand tracking has recently become an alternative input technology for the current generation of Augmented Reality (AR) and Virtual Reality (VR) Head-Mounted Displays (HMD). The devices sense the movements of the hands, including precise finger movements, and create models of the entire hand in real-time. In contrast to traditional hand-held controller input, the information we gain about the user is vastly increased. This information cannot only be used for input to extract the commands the user wants to put in but can also provide additional information about how the user performs the commands. This information is a valuable source for behavioral biometrics.

*Corresponding author: Jonathan Liebers. E-Mail: jonathan.liebers@uni-due.de



(a) Augmented Reality (AR). (b) Virtual Reality (VR). (c) Hand Tracking in AR and VR.

Figure 1.: We investigate the identification of users by their hand tracking data in (a) Augmented and (b) Virtual Reality. To do so, we collected hand tracking data from users interacting with frequently-used interface elements (such as buttons and sliders). The data is elicited by a head-mounted display that takes several different reference points of the user's hands into account (c). The reference points visible in (c) were not visible to users.

Behavioral biometrics in Augmented and Virtual Reality on HMDs facilitate implicit user identification. Implicit user identification enables the identification of a user through actions that the user would carry out anyway (Jakobsson, Shi, Golle, & Chow, 2009) in an implicit interaction (Schmidt, 2000). Implicit user identification allows for improvements in usability and user experience by customizing and personalizing the user interface to the preferences of the user in the case of a shared device. Further, it can be beneficial for granting access to personal information without the need to enter a password, as they are laborious to enter and open to attack, since an immersed Virtual Reality user can easily be observed by bystanders (George et al., 2017; Olade, Fleming, & Liang, 2020). Thus, recent works explored the use of hand-held controllers for implicit behavioral biometric user identification in Virtual Reality (Liebers, Abdelaziz, et al., 2021; Pfeuffer et al., 2019).

Since we currently observe a shift from controller-based to hand and finger-based interaction, we investigate how finger and hand tracking can be used to implicitly identify users. In particular, we focus on basic input elements used to create user interfaces in Augmented and Virtual Reality, such as buttons, sliders, and keyboards. In a user study conducted with 16 participants in two separate sessions on two separate days, we investigate how uniquely participants interacted with eight common input elements in Augmented and Virtual Reality. The study explores the influence of the device type (Augmented Reality vs. Virtual Reality), the number of hands used for input (unimanual vs. bimanual), and the user interface elements associated with four interactions (pointing vs. manipulating, see Aigner et al. (2012)) on user identification (cf., Figure 1). Overall, we were able to identify users over two days with up to 95% accuracy using an explainable machine learning approach. Upon further examination of the data, we also found that two-handed input outperformed one-handed input, particularly in Virtual Reality, and that Pointing Gestures performed better than Manipulation Gestures.

Contribution Statement. The contribution of this work is twofold. First, we create a prototype for implicit user identification using behavioral biometrics in AR and VR that is based on the users' hand tracking data, which is spatiotemporal data. We investigate the users' unimanual and bimanual interaction across two input gestures concerning eight different user interface elements and report on a user study that is spread across two sessions on different days ($N = 16$). Second, we evaluate which interactions benefit an identification system's identification accuracy and provide insights into our explainable machine learning classifier.

2. Related Work

User identification can generally be enabled in one of three ways: through knowledge-driven approaches, token-based approaches, or biometrics (Jain, Flynn, & Ross, 2007; Jain, Ross, & Nandakumar, 2011) In our case, however, we focus on behavioral biometrics. Behavioral biometrics is a subset of biometrics that focuses on the behavior of people (Jain et al., 2007). Since our approach makes use of standard HMDs for both AR and VR as well as the users' finger interactions, we focus on hand- and head-tracking-based approaches to biometric identification.

2.1. User Identification on Head-Mounted Displays

The current state of the art for user identification with HMDs involves mechanisms that are often derived from mobile phones, such as personal identification numbers (PINs), passwords, or pattern locks (Yu, Liang, Fleming, & Man, 2016). However, these traditional methods are susceptible to shoulder surfing attacks (George et al., 2017). George et al. (2017) find that an attacker who observes an immersed VR user's password entry is able to correctly perceive up to 18 % of all passwords that users enter with hand-held controllers . Olade, Liang, Fleming, and Champion (2020) confirmed this number, as they found that an attacker has a success rate of 20 % for pattern authentication in VR . Although this is quite a severe issue for VR, it is not exactly the same case for AR. In contrast to VR, the user is not as immersed in virtuality when wearing an AR-HMD; therefore, the user should be less susceptible to observation attacks, as they can still perceive reality and any potential attackers.

Nevertheless, these problems motivate further work in the field. Behavioral biometrics in particular have recently been investigated in this regard. Although it is not a direct consequence, behavioral biometrics can often be implemented in such a way that they are based on an implicit interaction (Schmidt, 2000). This implicit interaction forms an implicit identification based on actions that the user would carry out anyway (Jakobsson et al., 2009). Implicit identification not only relieves the user of any explicit interaction with the identification system, but it also can be utilized continuously in the background (Deutschmann, Nordström, & Nilsson, 2013; Schneegass, Oualil, & Bulling, 2016). A similar term was introduced by Corner and Noble (2002) in the context of token-based systems, denoted as "zero-interaction authentication". These implicit properties allow the frequent adjustment of authentication models to the ever-changing user behavior (Chauhan, Kwon, Hui, & Mascolo, 2020). These unique benefits strongly motivate the development of further approaches to identification for HMDs.

2.2. Biometrics in Augmented and Virtual Reality

For their unique benefits, an increasing number of biometric approaches were recently explored. Most previous work focused on Virtual Reality due to its popularity, but most principles are theoretically applicable to Augmented Reality HMDs as well. One unique benefit of head-mounted displays is the spatial relationship between the hand-held controllers and the worn HMD, which can be used for user authentication as explored by Pfeuffer et al. (2019). They state, in particular, that the users' body relations and distances as well as hand coordination are strong biometric features, where especially the dominant hand is of high importance (Pfeuffer et al., 2019). Our work ties in closely with their work for a number of reasons. First, they explored the general capabilities of implicit biometric identification in VR, which we further by including AR as well. Moreover, while they used hand-held controllers and ranged interactions (e. g., ranged pointing or ranged typing), we set our focus on near interactions that users execute with their hands and fingers, tracked by the HMD.

2.2.1. Task-driven Biometrics

Recently, new approaches that employ behavioral biometrics in relation to the execution of tasks, where the data elicited during the interaction is used for the purpose of identification, have been investigated. This data is often spatiotemporal data. For example, Kupin, Moeller, Jiang, Banerjee, and Banerjee (2019) showed that the task of throwing a ball in Virtual Reality elicits highly individual spatial motion data. Ajit, Banerjee, and Banerjee (2019) then conducted an extended analysis, yielding a higher accuracy on a larger data set. Similarly, Miller, Ajit, Kholgade Banerjee, and Banerjee (2019) created an extended solution for rejecting intruders and performed a within- and cross-system evaluation (Miller, Banerjee, & Banerjee, 2020), again with regard to the ball-throwing interaction. Later, Liebers, Abdelaziz, et al. (2021) explored the interactions of sports-based tasks, such as bowling and archery, in VR including a body normalization which increases identification accuracy across participants. In addition, Olade, Fleming, and Liang (2020) explored the task of grabbing, moving, and dropping balls into different containers for the purpose of identification in VR.

2.2.2. Identification through Interaction with User Interfaces

Some of the earliest examples of identification through user interfaces originate from keyboard (Monrose & Rubin, 2000) and mouse (Ahmed & Traore, 2007; Gamboa & Fred, 2004) biometrics, which were bound to desktop computers. Later, as touchscreens and mobile devices became prevalent, those core principles were transported to this new form of user interface (Buschek, De Luca, & Alt, 2015). However, it is not simply a new application of the older concepts. Touchscreens, for example, opened new possibilities for biometric authentication, such as using finger pressure as a biometric trait (Saevanee & Bhatarakosol, 2008). Currently, we are on the brink of again switching the most common user interfaces. While in earlier times the evolution from standard desktop computers to mobile devices already imposed a huge step, we are now on the threshold of moving to three-dimensional, virtual user interfaces, such as interface elements in VR, and holographic user interface elements. Thus, we see a strong need to explore these interactions with novel interface elements, understanding their possible association with behavioral biometrics.

2.3. Head-movement-based Biometrics

Miller, Herrera, Jun, Landay, and Bailenson (2020) recently explored the personal identifiability of user tracking data with over 500 participants, whose head movement was tracked as they observed 360° videos in VR. They found that they could identify

511 participants correctly at an accuracy of up to 95% with a system trained on less than five minutes of tracking data per person. Their number of participants exceeds the regularly reported numbers of participants by a great margin (Sugrim, Liu, & Lindqvist, 2019). With respect to external audio stimuli, Li et al. (2016) presented "Headbanger", a system they evaluated through experiments involving 95 participants on an AR device. In "VRCAuth", Sivasamy, Sastry, and Gopalan (2020) utilized two different datasets to explore the idea of continuous authentication of users by their spatial head movement data in VR. Alike, Mustafa, Matovu, Serwadda, and Muirhead (2018) conducted a user study that identifies people based on their head movement data in VR. Also, approaches exist that integrate user's head and neck modeling to implement a nodding interaction for unlocking their system (Wang & Zhang, 2021). With "MoveAR", Bhalla, Sluganovic, Krawiecka, and Martinovic (2021) presented an approach to continuous biometric authentication for AR headsets (HoloLens) that makes use of the spatial head movement data, whilst the participants performed certain hand gestures, such as typing. The head has also been shown to be important for gaze-based authentication (Liebers, Horn, Burschik, Gruenefeld, & Schneegass, 2021). HMDs have also been used to implement gait-based biometric systems, as shown by Shen et al. (2019), by utilizing the Google Glass.

2.4. Hand-tracking-based Biometrics

Before commercial off-the-shelf HMDs for Augmented and Virtual Reality were equipped with hand and finger tracking technology, the "Leap Motion"[1] was widely used to explore hand-related biometric features. The device senses hand gestures at an accuracy of 0.2mm for static and 1.2mm for dynamic setups (Weichert, Bachmann, Rudak, & Fisseler, 2013) and recognizes a hand model consisting of several bones via its sensors. Thereby, various systems for identification have been developed. One example comes from the work of Maruyama, Shin, Kim, and Chen (2017), who explored user authentication based on the hands' skeletal features, such as the distance between fingertips, as a biometric trait. Although their approach falls primarily into the domain of physiological biometrics, behavioral-biometrics-based approaches exist as well. Another example is the approach by Chan, Halevi, and Memon (2015) that also includes hand gestures as a form of behavior. However, in their discussion they state that hand geometry matters the most for identification. The Leap Motion has also been utilized to implement behavioral biometric authentication schemes, such as the handwriting in the air (Kamaishi & Uda, 2016). Similarly, Xiao, Milanova, and Xie (2016) developed a Leap-Motion-based behavioral signature verification. Ataş Musa (2017) further investigated a hand-tremor-based biometric recognition, testing whether hand tremor is unique for humans. It was positively evaluated in a user study with five subjects. Manabe and Yamana (2019) also developed a two-factor authentication system. They used a numeric keypad and a Leap Motion to determine identities through physiological features, such as the length of phalanges and metacarpals, but also by behavioral finger speed.

Another work related to hand and finger tracking in VR is "ElectricAuth" by Chen et al. (2021). In their work, the authors explored the feasibility of user authentication by externally actuating a user's fingers through electric muscle stimulation and capturing the output with an HMD. Although they capture finger movements, their

[1] Leap Motion by Ultraleap. `https://www.ultraleap.com/product/leap-motion-controller/`, last retrieved on October 20, 2022.

focus lies on the external actuation through electric muscle stimulation; hence, the participants' movements are induced by the system during the interaction. We regard their approach as a functional biometric (Liebers & Schneegass, 2020).

To our knowledge, there is no previous work that uses the hand tracking data of recent HMDs with interactions in Augmented and Virtual Reality for user identification.

3. General Approach

In this work, we investigate to which extent we can identify users in Augmented and Virtual Reality from their hand and finger-tracking-based interactions. Finger tracking is associated with a huge amount of data, as the virtual hand objects consist of various virtual bones (cf., Figure 1(c)). Each virtual hand object consists of 26 elements in total, where each element is defined by positional and rotational coordinates. For the different areas of the hand, we have the following number of elements: Thumb (4), Index-, Middle-, Ring-, Little-Finger (5), Palm (1), and Wrist (1). This complex nature seems very promising for creating a system that is able to distinguish users. Therefore, we investigate how well user identification works in Augmented and Virtual Reality concerning common user interface elements in two types of interactions: "pointing" and "manipulation" gestures, as defined by Aigner et al. (2012). We implemented these two gestures by four common user interface elements each and vary them for unimanual and bimanual interaction (cf., Figure 2). As the user interface elements are not only associated with a gesture but also with an interaction (e. g., how the arms need to be moved for the interaction), we group them into four sets: "Button Interaction" that contains button-press related user interface elements, "Keyboard Interaction" which are related to complex keyboard interactions, "Horizontal Manipulation" (e. g., the translation of an object on a horizontal axis) and "Diagonal Manipulation", the manipulation on a diagonal axis. In our design, we follow the dominant usages of the two gestures in previous work (Groenewald et al., 2016).

We created a software prototype that implements the user interface elements and the associated tasks. It elicits the data of the user's interactions which we then use for a post hoc analysis to determine the separability of the data by each individual person. We seek to use only one single code base that we can deploy to a head-mounted Augmented and Virtual Reality device respectively, to ensure that both technologies are as comparable as possible. For this reason, we opt for Microsoft's Mixed Reality Toolkit[2] (MRTK), which facilitates the creation of one single software prototype in Unity3D[3] that then can be deployed to our two target devices in a user study. Thereby we ensure that the user interface elements have the same positions, orientations, sizes, and behaviors across both devices. For devices, we chose the Microsoft HoloLens 2 as our AR device and the Meta Quest 2 as our VR device.

4. User Interface Elements for Interaction

This section describes our implementation of the three-dimensional user interface elements and the finger tracking in Augmented and Virtual Reality. All of the interac-

[2]Mixed Reality Toolkit by Microsoft. `https://github.com/microsoft/MixedRealityToolkit-Unity`, last access on October 20, 2022.

[3]Unity3D. `https://unity.com`, last access on October 20, 2022.

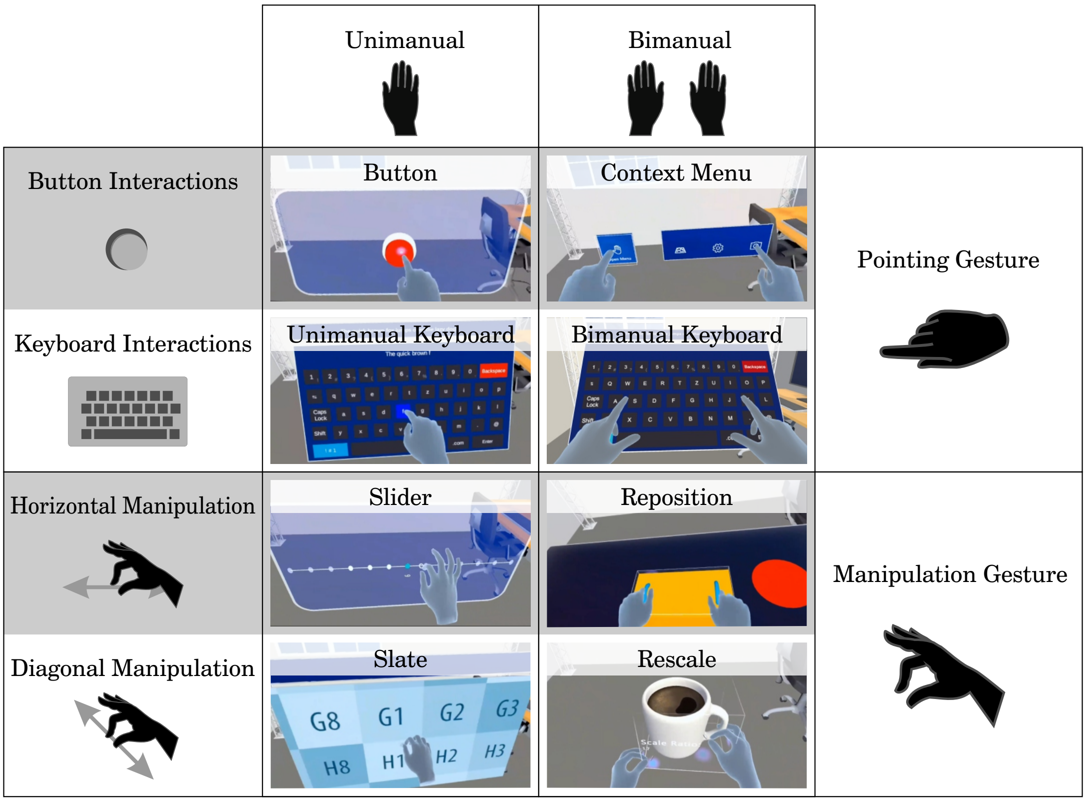


Figure 2.: The tasks and user interface elements that were implemented with the MRTK with regards to the variables i) number of involved hands (unimanual vs. bimanual) and ii) the user interface elements (Button Interaction vs. Keyboard Interaction vs. Horizontal Manipulation vs. Diagonal Manipulation). Button Interaction and Keyboard Interaction originate from a Pointing Gesture and Horizontal Manipulation and Diagonal Manipulation from a Manipulation Gesture.

tions, virtual objects, and hand representations were implemented in Unity3D using the Microsoft Mixed Reality Toolkit (MRTK). This combination allows the creation of one prototype that can be deployed to both of our target devices, the Meta Quest 2 and the Microsoft HoloLens 2, for an identical code-wise behavior of the prototype.

4.1. Hand Tracking

The users' hands are tracked by the hardware vendor's finger tracking model in Augmented and Virtual Reality. Both the HoloLens 2 and the Quest 2 devices employ a device-specific finger tracking recognition that maps the users' real hands into the virtual environment. Both devices track the hands by default at 30 Hz. The hand and finger tracking work in general through a machine learning model that uses the device's depth cameras (Han et al., 2020).

4.1.1. Hand Representation in Virtual and Augmented Reality

The users' hands are visually represented in Virtual Reality (cf., Figure 1(c)). Within the scope of this work, we used the default model that is provided by the MRTK to display the hands to the user, which renders the hands as opaque objects without any attached arms. All fingers are animated and the hand objects follow the position and rotation of the user's real hands.

In contrast to VR, the user can see their real hands in AR. Therefore, our work does not add a visual model as a representation of the users' hands. Still, the issues of visual and spatial registration arise here. The holograms that the HoloLens 2 draws are spatially registered at a certain point in 3D, but as they are rendered on the displays in front of the users' eyes, they overlap the hands visually. This effect happens because the user's gaze first falls through the display and, at a further point, to their hands. To enable correct visual registration, we added an occlusion shader to the users' hand model in AR. This results in the hologram not being rendered over the hand, thereby making the hands and the holograms correctly visually registered in 3D.

4.1.2. Press Cursor

To interact, the users' hands need to touch the virtual object. As this is sometimes hard to see, we added a "PokePointer" from the MRTK to the users' hands. A PokePointer is a small cursor denoted by a circle that is located near the index fingertip. It only appears when an interactable object is nearby.

4.2. Elements Design and Tasks

We designed a unimanual and a bimanual version of four different user interface elements (see Figure 3).

4.2.1. Button Interaction

The first user interface element is a red button (3.5×3.5 cm) that responds to single presses with the fingertips (cf., Figure 3(a)). For the unimanual input, the task was to press the button. For the bimanual input, we opted for a context menu. The context menu consists of a button and a group of three buttons, which is initially invisible and only appears once the other button is pressed by the user (cf. Figure 3(e)). If the user holds the first button with one hand, the menu appears on the right. The task was to first press the left-hand button to open the context menu and then to take a picture by pressing the camera button on the right hand side.

4.2.2. Keyboard Interaction

The keyboard has a size of 55 cm by 30 cm, is tilted by 30°, has a QWERTZ layout and is an adapted implementation of an asset from Unity's Assetstore[4] (cf., Figure 3(d)). We use this keyboard because the MRTK only provides access to the native system keyboard, which differs in size and layout between the Quest 2 and HoloLens 2 devices. We added a spell check to the keyboard which indicates a spelling mistake by switching the users' input to red color if the text entry is wrong. For the unimanual input, the keyboard only registers letters from one finger press at a time whereas, for the bimanual input, it supports the concurrent text entry of two fingers. The user has to enter the sentence "The quick brown fox jumps over the lazy dog" which is a pangram, meaning that it contains all letters of the English alphabet.

4.2.3. Horizontal Manipulation

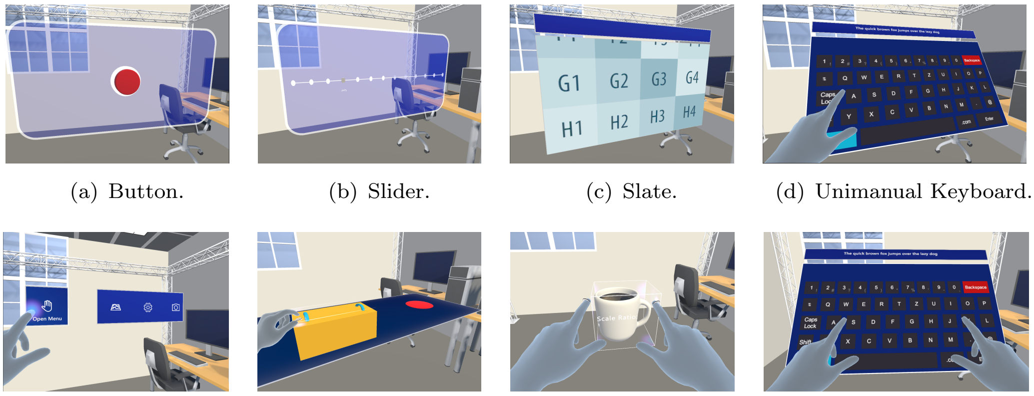
The slider used for Horizontal Manipulation allows the user to input a number from a fixed interval (cf., Figure 3(b)). It can take up to 13 different positions on a horizontal axis 5 cm apart from one another. The slider implements a discrete behavior. To complete this interaction, the slider has to be moved horizontally from position three to position nine. For the bimanual input, a yellow cube with a width of 30 cm, a height of 12 cm, and a depth of 15 cm needs to be repositioned by grabbing the edges of the object with both hands. The user's task was to move the yellow cube onto the red circular target area, which has a radius of 6 cm. Both objects are initially 70 cm apart. The cube indicates its functionality by a glowing wireframe near its edges that indicates a possible interaction. The interaction ends once the cube has been successfully moved to the red target area.

4.2.4. Diagonal Manipulation

The slate is a 2D pane that can be scrolled with a Diagonal Manipulation gesture by grabbing it with a pinch movement. It displays a checkered board that consists of cells, where each cell has its position imprinted. The participant's task was to slide the bottom left cell ("H1") into the upper right corner of the slate. The full movement covers a distance of 70 cm. For the bimanual input, we designed a rescaling task in which a coffee mug (10 cm x 12 cm x 10 cm) needs to be rescaled using both hands. The mug is accompanied by a floating label that indicates its current scale ratio as a decimal number. Its initial scale is "1" and needs to be changed by the users to "1.7" by pinching the mug with two hands at the top corners of the object. The corners are again indicated with a glowing wireframe that signalizes affordance to the rescaling motion.

---

[4]VR Keyboard. `https://assetstore.unity.com/packages/tools/input-management/vr-keyboard-95936`, last access on October 20, 2022.



(a) Button. (b) Slider. (c) Slate. (d) Unimanual Keyboard.

(e) Context Menu. (f) Reposition. (g) Rescale. (h) Bimanual keyboard.

Figure 3.: Overview of the user interface elements for unimanual (a-d) and bimanual (e-h) input. The respective user interface element of (f) and (g) are glowing wireframes that appear when a hand is near the object to signalize affordance.

4.3. Data collection

Data recording takes place exclusively on the device by logging the position (coordinates x, y, and z in the world space) and rotation (euler_x, euler_y, euler_z) of all objects in every rendered frame. These objects are the coordinates of the head, represented by the worn HMD, and the hands and the fingers of the user, as provided by the device's capabilities. The hands are represented by their anatomical features, i. e., each of the five fingers is also subdivided into its phalanges. For each phalange, we log its positional and rotational coordinates as well.

Furthermore, besides the spatial data, we also log the timestamp, a unique frame identifier, and an event column. The event column consists of all possible event callbacks provided by the MRTK (e.g., "ButtonPressStart" and "ButtonPressEnd" events for the button condition). The exact number and semantics of the events depend on the particular MRTK asset. For the keyboard, we log the length of the entered string as well as its value. Furthermore, once the interaction's end condition was met (e.g., the button was fully pressed and released), another marker was added to the event column. As the logging is bound to the application's frame rate, we obtain a sampling rate of 72 Hz for the Meta Quest 2 and 30 Hz for the Microsoft HoloLens 2.

5. Study

We verify our approach in a user study that took place in our lab across two sessions over different days. We evaluate the elicited spatiotemporal data within an identification system and also participants' quantitative feedback.

5.1. Study Design

The study followed a within-subjects repeated-measures design and it was split into two sessions that took place on two different days, sharing the same study design in both sessions. All participants participated in both sessions of the study. The study ran for a timeframe of three weeks. We chose this split for two reasons. First, the results obtained indicate that identification is possible across different days, as the user's behavior might subtly change in the meantime. Second, as the users had to take off the HMD between days, the identification model cannot determine the user's identity by having them wear the HMD in a certain "odd" way. The independent variables of this study are *Device* with two levels (Augmented Reality vs. Virtual Reality), and *Hands* with two levels (unimanual vs. bimanual), and *User Interface Elements* with four levels (Button Interaction vs. Keyboard Interaction vs. Horizontal Manipulation vs. Diagonal Manipulation).

5.2. Apparatus

We used two different devices in the study, one for AR and one for VR. For AR, we used the Microsoft HoloLens 2. The HoloLens 2 offers 6 Degrees-of-Freedom (DoF), a hologram refresh rate that we set to 30 Hz, and native support for hand- and finger-tracking. The diagonal field of view is 52° and it can operate wirelessly. For VR, we employed the Meta Quest 2 device. The Quest 2 also offers 6 DoF, and a variable display refresh rate that we set to 72 Hz for our study. Its display has a field of view of 95° and it can also operate wirelessly.

The study took place in our lab, which is a room with 5.1×3.1 m with a ceiling height of 3.1 m. Since the study part in AR takes place in the real lab environment, we modeled the same room as a virtual environment for VR so that the virtual environment shares the same dimensions and furniture (e.g., dimensions, tables, computers, and truss) alike the real laboratory (cf., Figure 1(a) and Figure 1(b)). Both devices execute the same Unity prototype application, with the important difference that the virtual room objects (e.g., the walls, furniture, and ceiling) are disabled in AR on the HoloLens 2, as they are provided by the real environment, where the user moves within.

To align the coordinate grids of the VR and AR HMD, we physically marked a spot in the laboratory as the origin point of the coordinate system and took care to initialize the HMDs at this specific point. Therefore, the coordinate grids for AR and VR are identical. Moreover, the experimenter set the safety guard for the Quest 2 in a way, that it aligns with the walls of the room. Due to the size of the room and the positioning of the stimuli, no participant had to move near the boundaries of the safety guard. Each of the user interface elements is located approximately 1.35 m above the ground level for ergonomics, measured from its center point. To the left-hand side of the user's initial position, we set up a virtual whiteboard that describes the current task within the study. After each task's completion, the application instructed the participant to move back to the origin point.

5.3. Procedure

The study took place under COVID-19 conditions, strictly following the local regulations. All devices and touched surfaces were cleaned between participants and the room was continuously aired. First, we welcomed the participants, explained the procedure and obtained written, informed consent from them, and explained that their participation in the study could be aborted at any time without any detriments. We also fully answered any questions that the participants had with regard to their participation in the study and the procedure. Further, we informed the participants that their movements and behavior would be recorded in the study by the means of the HMD (i. e., logging of any coordinate and in-app video recording) and as well by an external video camera. The participants obtained a short introduction in the form of a prerecorded video to the devices (Quest 2 and HoloLens 2). Then, the experimenter assisted the participants in wearing the device (i. e., adjusting the straps of the head-band and setting the inter-pupillary distance of the lens spacing on the Quest 2). For the HoloLens 2, the first action was to calibrate the device to the participant's eyes by launching the integrated calibration routine so that the projected holograms are clear to view.

After the initial setup, we tested two blocks, one for each device (AR and VR). Each block consists of the counterbalanced eight interactions with the User Interface Elements for unimanual and bimanual execution. The counterbalancing of the eight interactions as well as the order of the two blocks was performed using a Latin Square respectively. Once the desired number of repetitions was fulfilled per interaction, we verbally asked the participants for their rating of two Likert scale questions on a scale from 1 (strongly disagree) to 7 (strongly agree): i) "I found the interaction physically highly demanding" and ii) "I found the execution of the interaction highly natural". After each block, the participants took the HMD off and answered the NASA Raw Task-Load-Index (Raw TLX) questionnaire (Hart, 2006; Hart & Staveland, 1988), rating the workload of the interactions and the System Usability Scale (SUS) questionnaire (John Brooke, 1996), rating the usability of the interactions.

Each interaction with any user interface element could be tried out by the participants first, to familiarize themselves with the task. Then, each interaction with any user interface element but the uni- and the bimanual keyboard was repeated for 12 trials each. The uni- and bimanual keyboard was tested for only 6 trials, as the interaction is much more complex in comparison to the others. We counterbalanced the order of the blocks and the order of the user interface elements and excluded the uni- and bimanual keyboards from becoming a direct successor of each other. The procedure was repeated in a second session without the questionnaires. The first session took approximately 90 minutes to finish and the second session took about 60 minutes.

5.4. Ethics

The focus of this research lies on the identification of persons throughout their comprehensive spatial interaction data with virtual, three-dimensional user interface elements in an implicit interaction. If the insights from this work are applied to a real-world system, it is an absolute necessity that the users know about the utilization of the proposed principles in any application. They furthermore should express informed consent that their comprehensive interaction data is used in this particular way. We strongly recommend any future reuse of the presented research to account for the principles of user privacy in AR and VR (Adams et al., 2018; Pearlman, 2020). Biometric data collection is not a unique attribute of AR and VR applications, but the scope of the gathered information is on a new level in comparison to established, widespread devices (Dick, 2021). Furthermore, all local regulations must be taken into account at any time and proper precautions for data security should be met.

To conduct our user study, we followed the regulations of our institution and obtained clearance from the institute's ethics review.

5.5. Participants

We recruited 16 volunteers (6f, 10m, 0d) from our university's mailing list and social media who were between 19 and 33 years old (M=25.62, SD=3.56). All of the participants were right-handed. We also tested the participants for their dominant eye and obtained four times "left" and twelve times "right". Their mean height was 174.88cm (SD=9.04cm, min=158cm, max=190cm), as reported on their ID card. Moreover, we measured their arm length from the shoulder to the fingertips and obtained a result of 69.31cm on average (SD=4.36cm, min=60cm, max=76.5cm). Nine participants stated that they had a form of visual impairment and all but two were completely corrected during the study. In the other two cases, the participants confirmed that it would not interfere with the study. We asked the participants for their previous experience with VR on a 7-point Likert scale, where "1" corresponds to "I have never used VR before" and "7" corresponds to "I experience VR on a daily basis". The participants' response was a median value of 5 (IQR=5). We asked the same question with regards to their experience with AR and again obtained a median value of 5 (IQR=5).

5.6. Quantitative Results

Besides the spatial data that we employ within our identification system (cf., Section 6), we obtained several insights from our user study, including task completion times, Likert items, and the Raw TLX and SUS questionnaires.

5.6.1. Task Completion Times

We tracked participants' Task Completion Time (TCT) per each repetition of each condition in our study. To better understand the influences on the user's TCT that are imposed by the unimanual and bimanual interaction and the user interface elements, we aggregate the per-user TCT as the mean TCT of all repetitions of each interaction with any user interface element. We apply the Aligned Rank Transform (ART-C) procedure (Wobbrock, Findlater, Gergle, & Higgins, 2011) to the aggregated data before performing an analyses of variance (ANOVA) with the within-subject factors *Device* (AR vs. VR), *Hands* (unimanual vs. bimanual), and *User Interface Element* (Button Interaction vs. Keyboard Interaction vs. Horizontal Manipulation vs. Diagonal Manipulation).

We consider the effect of the two levels of *Device* (AR vs. VR) on the TCT. The mean times for the different levels of *Device* are (in descending order): AR = 12.23 s (SD: 17.37) and VR = 11.43 s (SD: 15.73). To investigate the effect of *Device* on TCT, we apply a one-way ANOVA on the aligned ranks, showing a significant effect between conditions $F(1, 225) = 26.05$, $p < .001$.

We also seek to understand the effect of *User Interface Elements* with four levels (Button Interaction vs. Keyboard Interaction vs. Horizontal Manipulation vs. Diagonal Manipulation) on the TCT. The mean times for the different levels of *User Interface Element* are (in descending order): Keyboard Interaction = 39.73 s (SD: 6.75), Horizontal Manipulation = 3.22 s (SD: 1.82), Diagonal Manipulation = 2.97 s (SD: 1.49) and Button Interaction = 1.43 s (SD: 1.33). We apply a one-way ANOVA on the aligned ranks and find a significant effect between the conditions $F(3, 225) = 5.93$, $p < .001$. Table 1 lists the results of the subsequent pairwise post hoc comparisons.

Furthermore, we also investigate the effect of the two levels of *Hands* (unimanual vs. bimanual) on the TCT. Here, the mean times for the different levels of *Hands*

Table 1.: Pairwise post hoc comparisons for investigating the interaction effect between the independent variables *Device* and *User Interface Elements* ("Inter."=interaction, "Manip."=manipulation and "M"=mean).

XR	Pairwise Comparison (a vs. b)	M(a)	M(b)	t	p
AR	*Button Inter.* vs. *Keyboard Inter.*	1.47	41.55	-21.37	$< .0001$
	Button Inter. vs. *Horizontal Manip.*	1.47	2.95	-9.23	$< .0001$
	Button Inter. vs. *Diagonal Manip.*	1.47	2.98	-8.65	$< .0001$
	Keyboard Inter. vs. *Horizontal Manip.*	41.55	2.95	12.14	$< .0001$
	Keyboard Inter. vs. *Diagonal Manip.*	41.55	2.98	12.72	$< .0001$
VR	*Button Inter.* vs. *Keyboard Inter.*	1.39	37.88	-21.03	$< .0001$
	Button Inter. vs. *Horizontal Manip.*	1.39	3.48	-10.28	$< .0001$
	Button Inter. vs. *Diagonal Manip.*	1.39	2.96	-8.82	$< .0001$
	Keyboard Inter. vs. *Horizontal Manip.*	37.88	3.48	10.75	$< .0001$
	Keyboard Inter. vs. *Diagonal Manip.*	37.88	2.96	12.22	$< .0001$

are (in descending order): Unimanual = 12.03 s (SD: 17.21) and Bimanual = 11.64 s (SD: 15.92). Therefore, we apply a one-way ANOVA on the aligned ranks, showing a significant effect between conditions $F(1, 225) = 7.76$, $p < .01$.

We consider the interaction effects between the three independent variables (*Device*, *Hands* and *User Interface Elements*). Here, we found an interaction effect between *Device* and *User Interface Elements* with a two-way ANOVA $F(3, 225) = 5.93$, $p < .001$ on the aligned ranks. Moreover, we found another significant interaction effect between *User Interface Elements* and *Hands* with a two-way ANOVA $F(3, 225) = 22.89$, $p < .001$.

5.6.2. Individual Likert-Items

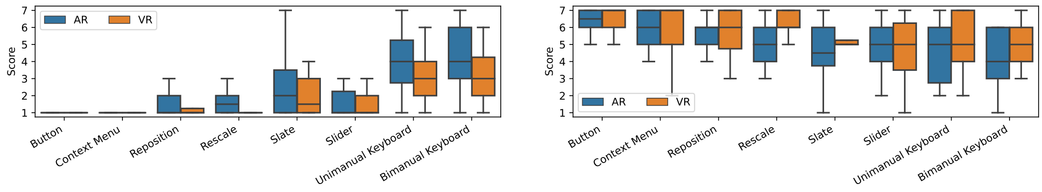
After completing any task during the study (e. g., after finishing the twelve button repetitions), we asked the participants two questions: i) "I found the interaction physically highly demanding" and ii) "I found the interaction highly natural". Both questions ranged on a Likert scale from 1 to 7, where "1" stands for "I strongly disagree" and "7" stands for "I strongly agree". The results are displayed in Figure 4.

We create a per-participant mean of all given responses and compare AR against VR using paired Wilcoxon Signed-rank tests. We were not able to find significant differences in the participants' responses.

5.6.3. Standardized Questionnaires

After the participants finished either the Augmented or Virtual Reality block, we asked them to answer the System Usability Scale questionnaire (John Brooke, 1996) and the NASA Raw Task Load Index questionnaire (Hart, 2006; Hart & Staveland, 1988). The results are illustrated in Figure 5.

5.6.3.1. Task Load Index. The participant responses to the NASA Raw-TLX are shown in Figure 5(b). The resulting score for AR (Md=38.75, IQR=20.63) is significantly higher than for VR (Md=31.67, IQR=18.33) with $W = 106.5$, $Z = 1.99$, $p = .049$, indicating lower workload for VR than AR through a Wilcoxon signed-rank



(a) "I found the interaction physically highly demanding."

(b) "I found the interaction highly natural."

Figure 4.: Participants' ratings of each interaction to the given questions in (a) and (b) on a Likert scale of 1 ("strongly disagree") to 7 ("strongly agree"). Outliers are not displayed.

test.

5.6.3.2. System Usability Scale. The participants' responses to the individual SUS items are illustrated in Figure 5(a). The median SUS score across all participants for AR is 63.75 (IQR: 63.75) and its mean value is 60.94 (SD: 19.30). For VR, we determined a median SUS score of 77.50 (IQR: 77.50) and a mean of 78.91 (SD: 12.75). We then compare the acquired SUS scores for AR against the SUS scores for VR using a paired Wilcoxon Signed-rank test and find a significant difference with $W = 4$, $Z = -2.6835$, $p = .0073$, $r = .4744$.

Bangor, Kortum, and Miller (2009) state that they determined a mean SUS score of 69.5 over 273 previous works which denotes an acceptable usability. We also compare our acquired SUS scores for Augmented and Virtual Reality against this hypothetical value using a Wilcoxon test (AR: $W = 9, p = 0.9184$ and VR: $W = 3, p = 0.3574$), but were unable to find a significant difference here.

6. Analysis of Tracking Data

To determine the identities of our users, we use a closed-set Random Forest multi-class classifier. This classifier trains on the data from the first session of our study and predicts the identities based on the data elicited during the second session.The predictions of the second day were only used for validation, showing that reidentification is possible across days.

6.1. Preprocessing and Feature Sets

To preprocess our elicited data from the study, we first split the data stream by interaction. Any interaction has a defined beginning and end (e. g., the button being fully pressed) and we can split the stream of data into segments every time each repetition (i. e., each study trial) ends. Next, we want to make our elicited data invariant to the global coordinate frame. Therefore, we subtract all rotational coordinates ("rot.x", "rot.y" and "rot.z") and the positional "pos.x" and "pos.z"-coordinates from the respective initial coordinates and only leave the "pos.y"-coordinate intact, which corresponds to the level of the HMD above ground level (i. e., the user's height). At last, we apply a modulo 360 function to all rotational coordinates, as they otherwise

Table 2.: Table of all evaluated Feature Sets with their descriptions and cardinalities (Card.). Any object consists of three positional coordinates ("pos.x", "pos.y" and "pos.z") and three rotational Euler angles ("rot.x", "rot.y" and "rot.z"). The abbreviations "I" and "T" stand for the Index Finger and Thumb respectively, "O" stands for "Other Fingers" (i. e., Middle Finger, Ring Finger and Little Finger). The subscript "Tip" indicates, that only the fingertip is taken into account, while "All" denotes that all finger-related objects of the model are considered (e. g., phalanges), including the fingertip.

F.-Set	**F0**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**	**F11**	**F12**
Card. Unim.	6	6	24	12	42	30	126	12	30	18	48	36	132
Card. Bim.	6	12	48	24	84	60	252	18	54	30	90	66	258
HMD	×							×	×	×	×	×	×
I_{Tip}		×	×	×	×	×	×	×	×	×	×	×	×
I_{All}			×		×		×		×		×		×
T_{Tip}				×	×	×	×			×	×	×	×
T_{All}					×		×				×		×
O_{Tip}						×	×					×	×
O_{All}							×						×
$Palm\ \&\ Wrist$							×						×

would possibly contain negative values. This approach makes the coordinates of the users invariant to the global coordinate frame so that we do not classify their change in position concerning their origin or have similar side effects.

In the next step, we take these segments and calculate the mean, min, max, and standard deviation of all given features. As a result, we obtain four floating point values that represent each feature, forming our feature vector.

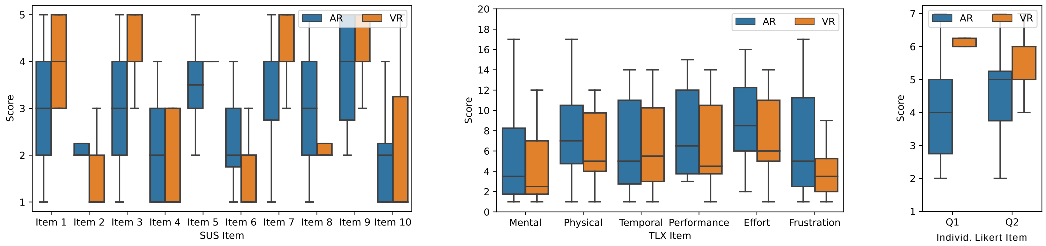
We implemented our preprocessing for two reasons. First of all, we followed the work of Pfeuffer et al. (2019), who chose a similar preprocessing and feature transformation. Second, by generating such feature vectors we obtain a flattened vector which in turn has the same shape for every elicited sample. The unified shape is necessary for the classifier. The feature vector consists of a variable amount of features, depending on the selected Feature Sets. We test multiple combinations by selecting different features from the available set of objects. In general, each hand consists of 26 objects and each object is represented by three positional coordinates ("pos.x", "pos.y" and "pos.z") and three rotational Euler coordinates ("rot.x", "rot.y" and "rot.z"). For bimanual interactions, we always include both hands, while for unimanual interactions, we only include the dominant hand.

When a hand is outside the sensors' field of view of the device, it cannot be tracked. Hence, we set the tracked values to NaN (not a number). Our mean, min, max, and standard deviation aggregate functions ignore NaN values.

We then select multiple Feature Sets from the overall given data, according to Table 2.

6.2. Model Definition

We chose a Random Forest classifier as our classifier for two reasons. First, it is in line with the approach of Pfeuffer et al. (2019) and second, the Random Forest is able to explain the features it uses for its classification predictions. We use the default



(a) System Usability Scale (b) Task Load Index (c) Likert Items

Figure 5.: Participants' ratings of (a) the System Usability Score (SUS), (b) the Raw NASA Task Load Index (TLX), and (c) the individual Likert items "Q1" and "Q2". Outliers are not displayed. Please note that higher scores do not always indicate better ratings.

hyperparameters by scikit-learn (Pedregosa et al., 2011) but set the "n_estimators" variable to 5000 and train it with the feature vectors that we obtained by aggregating each elicited feature from the study with the mean, min, max, and standard deviation aggregation functions.

We train one Random Forest per Feature Set, per user interface element, and per Augmented and Virtual Reality device with the data obtained during the associated study condition from the first session (i. e., one Random Forest was trained per entry in Table 3). We keep the study data strictly separated by these conditions.

We evaluate the Random Forest based on its reported accuracy values. As the amount of data per class and the amount of data per session is strictly balanced, no skew of accuracy is introduced. Alike other conventional Random Forest models, ours is a "closed-set" classifier, which means that it predicts one class out of a given set of known classes, where the classes correspond to user identities in our case.

6.3. Model Evaluation Results

In this section, we report the identification results of our classifier, i. e., the accuracy of identifying individuals by the given hand tracking data. Moreover, we discuss the features most relevant to the Random Forest model.

6.3.1. Identification Results

The overall identification results of our Random Forest classifier are listed in Table 3. Each accuracy was calculated by validating the Random Forest classifier that was trained with data from the first session of the study with data from the second session of the study per Feature Set.

The Feature Set F0 acts as the baseline Feature Set that only consists of the head. It achieves a mean accuracy of 0.55 for AR and 0.52 for VR. We identify F6 as the highest scoring Feature Set in the group of non-head including Feature Sets (F1-F6). At mean accuracies of 0.53 for AR and 0.58 for VR, it offers the highest accuracy at 0.88 again for the bimanual keyboard in AR and 0.83 in VR. It contains all the virtual bones and fingertips of the hand. F10 yields the highest mean accuracy for the Feature Sets including hand and head data (F7-F12). It results in an accuracy of 0.63 for AR and 0.64 for VR. Moreover, it contains the interaction that was classified with the

Table 3.: Overview of all acquired accuracies for all Feature Sets per user interface element in Augmented and Virtual Reality, accompanied by a mean accuracy over all eight interactions per Augmented and Virtual Reality. The highest scoring Feature Sets, determined by their mean accuracy of Augmented and Virtual Reality, are marked in bold. "C.-Menu" stands for the Context Menu interaction, "Uni.-Keyb." and "Bim.-Keyb." for the unimanual and bimanual keyboard respectively. The probability of a random guess is $1/16 = 0.0625$ and a darker cell shading indicates a higher value.

	Button		**C.-Menu**		**Reposition**		**Rescale**		**Slate**		**Slider**		**Uni.-Keyb.**		**Bim.-Keyb.**		**Mean acc.**	
F.-Set	AR	VR	AR	VR	AR	VR	AR	VR	AR	VR	AR	VR	AR	VR	AR	VR	AR	VR
F0	0.54	0.52	0.45	0.36	0.55	0.60	0.42	0.59	0.69	0.42	0.44	0.46	0.58	0.64	0.72	0.55	0.55	0.52
F1	0.35	0.45	0.46	0.51	0.42	0.49	0.47	0.51	0.43	0.39	0.33	0.43	0.34	0.40	0.80	0.73	0.45	0.49
F2	0.43	0.49	0.43	0.55	0.46	0.49	0.50	0.55	0.44	0.43	0.47	0.41	0.43	0.44	0.89	0.75	0.51	0.51
F3	0.51	0.53	0.43	0.57	0.45	0.53	0.49	0.58	0.51	0.43	0.33	0.45	0.50	0.57	0.82	0.84	0.51	0.56
F4	0.47	0.44	0.44	0.49	0.40	0.61	0.44	0.57	0.47	0.37	0.37	0.48	0.44	0.50	0.78	0.78	0.48	0.53
F5	0.45	0.48	0.51	0.60	0.43	0.58	0.47	0.57	0.52	0.41	0.35	0.46	0.65	0.54	0.76	0.76	0.52	0.55
F6	0.48	0.49	0.46	0.64	0.42	0.60	0.47	0.60	0.56	0.42	0.34	0.48	0.65	0.55	0.88	0.83	**0.53**	**0.58**
F7	0.57	0.62	0.47	0.56	0.49	0.59	0.57	0.66	0.72	0.48	0.53	0.49	0.70	0.62	0.93	0.78	0.62	0.60
F8	0.55	0.64	0.45	0.64	0.48	0.64	0.58	0.65	0.70	0.51	0.53	0.53	0.58	0.51	0.95	0.82	0.60	0.62
F9	0.49	0.45	0.46	0.57	0.43	0.61	0.48	0.59	0.59	0.52	0.31	0.56	0.58	0.54	0.83	0.78	0.52	0.58
F10	0.63	0.61	0.47	0.64	0.48	0.63	0.61	0.66	0.70	0.50	0.51	0.53	0.72	0.64	0.95	0.88	**0.63**	**0.64**
F11	0.44	0.46	0.48	0.49	0.41	0.64	0.39	0.59	0.61	0.48	0.30	0.45	0.59	0.58	0.81	0.79	0.50	0.56
F12	0.54	0.61	0.47	0.69	0.43	0.63	0.48	0.63	0.66	0.50	0.42	0.57	0.73	0.64	0.86	0.83	0.57	0.64

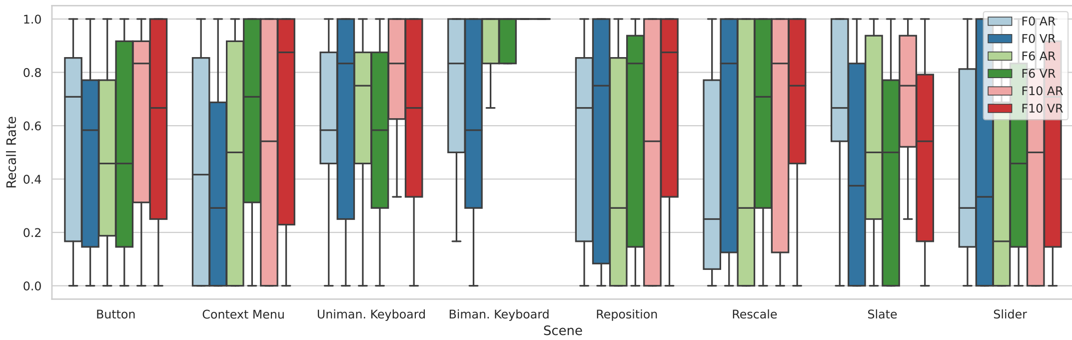


Figure 6.: Per-participant recall rates obtained from the Random Forest classifier for all User Interface Elements. The colors indicate values obtained from the Feature Sets F0, F6 and F10.

highest single accuracy for Augmented and Virtual Reality, the bimanual keyboard (0.95). F10 consists of the head and all virtual bones of the index finger and thumb, including the fingertip.

6.3.2. Feature Importances

We take a closer look at the most important features that the Random Forest classifier uses to distinguish between participants for the highest scoring Feature Sets F6 and F10. We do so, by examining scikit-learn's Random Forest through the "permutation_importance"-function with ten repetitions, using permutation feature importance. The permutation performs on the basis of our feature vector that consists of the values obtained from the aggregate functions mean, minimum, maximum, and standard deviation for each feature in our recorded time series. As the cardinalities of the Feature Sets are extensive, we discuss only the ten most important features that we obtain from the Random Forest and since the aggregate functions in the feature vector are hard to interpret on their own, we calculate a feature-wise mean value over all four aggregate functions and all user interface elements respectively. We then create a descending ranking of feature importance and focus on the ten best features.

We look into the most important features for the unimanual and bimanual interaction for Feature Set F6 and find that components of all fingers are equally represented. For the interactions with a Pointing Gesture, we find that they only consist of rotational coordinates, as all positional coordinates are absent within the ten most important features for the unimanual interaction and also for the ten most important features for the bimanual interaction. We furthermore see, that in these 20 features, a "rot.x" feature of the fingers is encountered 9 in 20 times. As the "rot.x"-axis corresponds to the "roll" of the hand, we believe that this is an important behavioral aspect. All 11 other features consist of "rot.y" and "rot.z" components that denote the yaw and pitch respectively.

For the Manipulation Gesture in F6, uni- and bimanual, we find that these as well consist only of rotational features, but "rot.x" exists only in 2 out of 20 entries. All the other features consist of "rot.y" and "rot.z", therefore, for these interactions we mainly see an individual change in the pitch and yaw of the hands. For bimanual interactions, the finger elements of the non-dominant hand are equally represented as for the dominant one.

For the Feature Set F10 the highest rated two most important features are head-related, in particular, "Head.pos.y" and "Head.rot.z". "Head.pos.y" directly symbolizes the height of the HMD above the ground, i. e., the participants' height. Also, "Head.rot.z" is the pitch of the head, i. e., whether they tilted their head down or up. Both features also are directly related. Since our user interface elements were always at the same height above the ground, a taller person had to tilt their head further down, explaining, why both features are important. Yet, it needs to be mentioned that an important feature does not necessarily yield an increased predictive accuracy (Breiman, 2001). This is encountered for both, unimanual and bimanual interactions in F10. All other features are finger related.

6.3.3. Effects on Recall Rate

To better understand the influences on the classifier's accuracy that are imposed by the independent variables, *Device* with two levels (Augmented vs. Virtual Reality) *Hands* with two levels (unimanual vs. bimanual interaction) and the *User Interface Elements* with four levels (Button Interaction vs. Keyboard Interaction vs. Horizontal Manipulation vs. Diagonal Manipulation), we aggregate the per participant's recall values on which the accuracies reported in Table 3 are based on (see Figure 6). We again apply the Aligned Rank Transform (ART-C) procedure (Wobbrock et al., 2011) to the aggregated data before performing an analysis of variance (ANOVA) on the aligned ranks. For pairwise post hoc comparisons, we used Holm-Bonferroni corrected t-tests. In the following, we report only the significant results. We do so, for three Feature Sets: F0, which acts as a baseline, including only the head, and F6 and F10, which are the highest scoring ones containing hand tracking data, including and excluding the head respectively.

6.3.3.1. Feature Set F0. We consider the interaction effect between between *Device*, *User Interface Elements* and *Hands*. The mean recall rates for the different levels of *Device* in descending order are 0.55 (SD: 0.37) for Augmented Reality and 0.52 (SD:

0.41) for Virtual Reality. For the four different levels of *User Interface Elements*, the mean recall values are 0.47 (SD: 0.38) for Button Interaction, 0.62 (SD: 0.37) for Keyboard Interaction, 0.51 (SD: 0.41) for Horizontal Manipulation and 0.53 (SD: 0.41) for Diagonal Manipulation. For the two different levels of *Hands*, we obtain a mean recall value of 0.54 (SD: 0.39) for unimanual interaction and 0.53 (SD: 0.39) for bimanual interaction. We apply a three-way ANOVA on the aligned ranks and find a significant interaction effect between the conditions ($F(3, 225) = 3.04, p = .03$), but any post hoc pairwise comparison did not reveal a statistically significant difference.

6.3.3.2. Feature Set F6. Furthermore, we also investigate the effect of the two levels of *Hands* (unimanual vs. bimanual) on the Recall rate in F6, which is the best Feature Set excluding the head object. The mean recall values are (in descending order): 0.50 (SD: 0.38) for unimanual and 0.61 (SD: 0.41) for bimanual interactions. We then investigate the effect of *Hands* on the recall rate by applying a one-way ANOVA on the aligned ranks, showing a significant effect between conditions ($F(1, 225) = 5.96), p = 0.0153$). We found that bimanual interactions ($M = 0.61$, $SD = 0.41$) provide significantly higher recall rates than unimanual interactions ($M = 0.50$, $SD = 0.38$), $F(1, 225) = 5.96$, $p = .0153$). We also consider the effect of the four levels of *User Interface Elements* (Button Interaction vs. Keyboard Interaction vs. Horizontal Manipulation vs. Diagonal Manipulation) on the recall rate. The mean recall rates for the different levels of *User Interface Elements* are: 0.52 (SD=0.39) for Button Interaction, 0.73 (SD=0.33) for Keyboard Interaction, 0.46 (SD=0.41) for Horizontal Manipulation and 0.52 (SD=0.40) for Diagonal Manipulation. A one-way ANOVA on the aligned ranks finds a significant effect within the *User Interface Elements* conditions, $F(1, 225) = 5.28, p = .0015$. The pairwise comparison found that Keyboard Interaction produced statistically significant higher recall rates compared to Button Interaction, $t(225) = -2.76$, $p = .0318$, Horizontal Manipulation, $t(225) = 3.816$, $p = .0011$, and Diagonal Manipulation, $t(225) = 2.65$, $p = .0344$.

6.3.3.3. Feature Set F10. We found statistically significant differences within the *User Interface Elements*, $F(3, 225) = 4.80$, $p = .0029$. The mean times for the different levels of *User Interface Elements* are (in descending order): 0.59 (SD: 0.41) for Button Interaction, 0.79 (SD: 0.33) for Keyboard Interaction, 0.54 (SD: 0.41) for Horizontal Manipulation and 0.62 (SD: 0.37) for Diagonal Manipulation. The pairwise comparison yields significant differences for Button Interaction vs. Keyboard Interaction, $t(225) = -2.69$, $p = .0390$, and Keyboard Interaction vs. Horizontal Manipulation, $t(225) = 45.72$, $p = .0019$.

6.3.4. Identification Accuracy during shorter Keyboard Interactions

Next, we investigate how the high Task Completion Time for the keyboard interaction influenced the identification accuracy. To do so, we re-evaluate the Random Forest for the unimanual and bimanual keyboard tasks with shorter excerpts of the interactions, where we perform the re-evaluation with a subsequently increasing amount of entered key presses. First, we re-train the Random Forest only with the very first key press and in the next step with the first and second key presses. Then we proceed with the first, second, and third key presses and repeat this procedure until we meet 43 key presses which is the desired input length of the pangram. We then take the results of all four Random Forests (Augmented and Virtual Reality $\times$ Unimanual and Bimanual

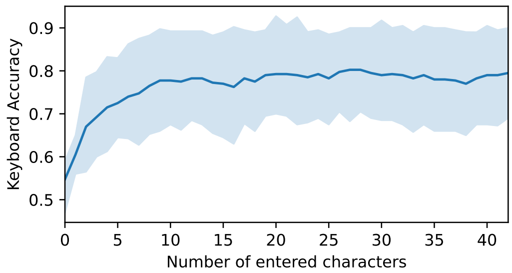


Figure 7.: Identification accuracy of the unimanual/bimanual keyboard in AR and VR over the number of entered characters.

Interaction) and display the results in Figure 7.

7. Discussion

In this section, we discuss our findings. We mainly focus on aspects that are important for user identification but compare Augmented and Virtual Reality as well.

User identification based on head-mounted displays and hand tracking is promising. We see from our results, in particular from Feature Set 10, that user identification based on hand tracking data is possible to a large extent. The best method is the bimanual keyboard, which performs better in AR (95% accuracy) than in VR (88% accuracy; cf., Table 3). We find for Feature Set 6 that the accuracy of the keyboard is significantly higher than any other user interface element. For Feature Set 10, the keyboard's accuracy is significantly higher than the interface elements associated with Button Interaction and Horizontal Manipulation. Here, we were unable to show a significant increase over Diagonal Manipulation that follows the keyboard closely in Table 3.

Bimanual interactions yield more individual data than unimanual interactions. Although this appears to naturally be the case, as it is expected for two hands to provide more data than one hand, we can also draw this conclusion from the data of Table 3. Here, we see a mean increase of 10% identification accuracy for the bimanual interactions compared to unimanual. For F6, which excludes the head, this poses a significant increase in the recall rates.

The head is a strong biometric feature. The head, as tracked by the HMD, is a strong biometric feature as suggested by the literature (cf., Section 2.3). We can see this from the increase in identification rate in F10 and also at the strong baseline that is imposed by F0. Moreover, for F10, the Random Forest classifier places the head-related positional features amongst the most important ones. Here, in particular, the users' height, as denoted by the "pos.y"-coordinate, has a major influence. Nevertheless, the hands perform similarly well, as can be seen by the mean accuracy of F6 vs. F0 (cf., Table 3). However, to acquire the highest possible identification rate, the head is a crucial element, as suggested by previous literature as well (Liebers, Horn, et al., 2021; Mustafa et al., 2018; Sivasamy et al., 2020).

The performance of the Keyboard is time-related. Keyboards are in general associated with highly individual behavior during the interaction, as previous works suggest (Buschek et al., 2015; Monrose & Rubin, 2000). With an accuracy of up to 95% our results yield similar performance as presented in related work for traditional and virtual keyboards. The performance of the keyboard in our research outshines the performance of any other user interface element significantly, except Diagonal Manipulation for F10, as we were unable to find a significant effect there. However, this high accuracy is also related to the significantly longer interaction times. The high performance of the keyboard is mainly met after 10 key presses. We can see this effect when we reevaluate our keyboard classifiers (cf., Figure 7). Therefore, we would argue that approximately ten seconds of keyboard interaction is enough to acquire an almost maximized identification rate. This rate can, of course, be increased when a more elaborate feature engineering is taken into account.

Hand-based interactions yield more individual data than hand-held controllers. Overall, we took a similar approach as Pfeuffer et al. (2019) in terms of study design, pre- and feature processing and classification. In their work, they implemented a ranged pointing task, a near-interaction grabbing task, a walking task, and a ranged typing task that were performed by 19 participants through hand-held controllers across two study sessions. In general, our work is most comparable with their grabbing task, as all the other tasks were performed from a distance with hand-held controllers, while we only use near-interactions with the user's hands. Here, one large difference lies within the hand-held controller providing fewer features, as the controller provides only one positional and rotational set of coordinates. However, the hand is composited of various virtual bones, which results in a multitude of captured values.

We saw for our data that if we exclude the head (F6), mainly the rotational coordinates of the hands and fingers are associated with high importance by the Random Forest classifier. Here, the positional coordinates play almost no role in the top-ten most important features. This is interesting, as hands can provide more rotational coordinates in comparison to hand-held controllers, but the rotation appears to be more important than the position. This certainly opens up a new dimension, in particular regarding previous findings of Pfeuffer et al. (2019). As in their work, we form similar feature vectors and the Random Forest classifier differs only in one hyperparameter (i.e., the number of estimators: 100 vs. 5000). We chose this larger number because our largest feature vector has a size of 1032 aggregated features. For increased explainability, we use Euler angles, whereas they use Quaternion angles. Although our base chance is 0.99% higher compared to theirs (Pfeuffer et al., 2019), we generally see an increased accuracy in our work. Their optimized result for the "grabbing" task is reported as 45.84%, which is surpassed by all accuracies that we determine in F10. For F6, which excludes the head, only the reposition and slider tasks in AR and the slate task in VR yielded lower accuracies. We, therefore, believe that the hands as an input modality yield much more unique data, which enables a better distinguishment between people, than hand-held controllers.

Capabilities of our Behavioral Biometric Model and Approach. We evaluated our Random Forest classifier on the data set that we elicited during our user study and the acquired accuracy values within a range of 30% (Slider in AR for F11) to 95% (Bimanual Keyboard in AR for F8 and F10), depending on the task, device and Feature Set. The knowledge-driven authentication method of entering a password is connected to a failure rate of 10% (Brostoff & Sasse, 2003). Although these values are not exactly comparable (e. g., due to differences in input modality, sample size, and a slightly different underlying problem), it still shows that in many cases the employment of a password-based authentication scheme exceeds the capabilities of our model. However, this is expected for several reasons. First, our approach focuses on the identification problem instead of authentication through verification (Jain et al., 2007). Thereby, our model yields a predicted identity as output instead of a binary "accept" or "reject". Identification is in comparison more convenient for users, as otherwise, users have to claim an identity beforehand, such as by selecting an account they want to authenticate with (Jain et al., 2007). Second, the proposed scheme is intended to be used in an implicit interaction and thus is suitable for "continuous identification" (Jakobsson et al., 2009; Traoré & Ahmed, 2012). At last, it also needs to be kept in mind that entering a password takes up considerable time in VR as George et al. (2017) for example state a mean value of 2.95 s for a four-digit PIN entry with controllers, whereas our interactions are intended to be part of an implicit identification scheme, where the determined interaction is a byproduct of regular interaction with the user interface. It thereby does not explicitly demand time from the user. The implicit employment within a continuous scheme allows for frequent re-identification of users that does not bother the user and can be used in a probabilistic scheme, which can also be associated with an increase in security (Traoré & Ahmed, 2012). Biometrics, in general, can also be used as a complement to passwords instead of a replacement (Hamid, 2015), which can still be requested as a fallback authentication method. Also, we believe that the employment of a deep learning model could increase the accuracy beyond the capabilities of the Random Forest in future work, hence we release the data elicited in the study to the public.

Augmented Reality vs. Virtual Reality. From our participants' feedback to our questions and questionnaires, we see that VR is associated with significantly higher usability, as denoted by the SUS, in comparison to AR. Although we created one code base and deployed the same application to both AR and VR devices, participants strongly preferred the interactions and hand tracking in VR.

Observations from the study show that participants' gestures were more often misrecognized in AR. Although our prototype for AR was rated below an acceptable usability score, we were unable to find a significant effect in this rating. Judging from a standpoint of user identification, AR on Head-mounted displays is on approximately the same level of identification accuracy as VR, even though the accuracy values vary minimally per condition, most likely due to individual noise. We furthermore were unable to find significant differences between the acquired recall rates of AR and VR.

Stability of the approach. The "stability" of a biometric approach refers to the fact that the biometric maintains its distinctiveness over time (O'Gorman, 2003). Our user study and the associated split of data, where we trained with the first session and validated with the second session, proves that the human spatiotemporal behavior during the interaction with the user interface elements retains its uniqueness over a short period of time. Longer periods were, to the best of our knowledge, not yet extensively covered in research with the exception of Miller, Banerjee, and Banerjee (2022) who recently reported on an investigation in the order of 7 to 18 months for the task of throwing a ball in VR.

8. Conclusion

In this work, we show that implicit identification of users by their spatial motion data elicited from their hands is possible in AR and VR with an accuracy of up to 95%. We evaluate in total eight different universal user interface elements that respond to nearby finger interactions in 3D and see that the keyboard in particular stands out. Still, we find that the keyboard's high accuracy rating is primarily associated with a high interaction time. Besides the keyboard, Diagonal Manipulation also appears to be very feasible for user identification. We show and evaluate these performances for

16 participants, whose data was elicited in a lab study that took place in two sessions on different days. Furthermore, we provide insight into our explainable machine learning approach (Random Forest) and show that the rotational coordinates of the hands are especially important. However, we also find that the head on its own is a strong biometric. Although AR is associated with a number of issues, we also see that the identification rate is only slightly impacted. Therefore, we believe that both technologies are very suitable for eliciting highly individual data from the hands and fingers.

9. Data availability statement

The data set elicited in the user study that supports the findings is openly available online at: `https://identifying-users-by-hand-tracking-data.hcigroup.de`

10. Acknowledgements

This work was supported by the German Research Foundation (DFG) under Grant Nr. 426052422.

About the Authors

Jonathan Liebers obtained a bachelor's and master's degree in computer science at the University of Duisburg-Essen. He currently is a Ph.D. student at the Human-Computer Interaction Group, University of Duisburg-Essen. His research interest lies mainly within the field of usable security and implicit identification.

Sascha Brockel obtained a bachelor's and master's degree in business information systems during his studies at the University of Duisburg-Essen from 2016 to 2021. He currently works in the field of process automation in the context of organizational and business management at KPMG. His interest encompasses automation, IoT, and reliable identity management.

Uwe Gruenefeld is a postdoc researcher in human-computer interaction at the University of Duisburg-Essen, Germany. He is fascinated by a wide range of topics around AR and VR. His research has mainly focused on visualizing out-of-view objects, passive haptics, usable security, and cross-reality systems.

Stefan Schneegass is an assistant professor of human-computer interaction at the University of Duisburg-Essen. His research interests include the crossroads of human-computer interaction and ubiquitous computing, particularly the development of implicit authentication mechanisms. Schneegass received a Ph.D. in computer science from the University of Stuttgart, Germany.

References

Adams, D., Bah, A., Barwulor, C., Musaby, N., Pitkin, K., & Redmiles, E. M. (2018, August). Ethics emerging: the story of privacy and security perceptions in virtual reality. In *Fourteenth symposium on usable privacy and security (SOUPS 2018)* (pp. 427–442). Baltimore, MD: USENIX Association. Retrieved from `https://www.usenix.org/conference/soups2018/presentation/adams`

Ahmed, A. A. E., & Traore, I. (2007, July). A new biometric technology based on mouse dynamics. *IEEE Transactions on Dependable and Secure Computing*, *4*(3), 165-179. Retrieved from `https://dx.doi.org/10.1109/TDSC.2007.70207`

Aigner, R., Wigdor, D., Benko, H., Haller, M., Lindbauer, D., Ion, A., ... Koh, J. (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. *Microsoft Research TechReport MSR-TR-2012-111*, *2*, 30.

Ajit, A., Banerjee, N., & Banerjee, S. (2019, dec). Combining pairwise feature matches from device trajectories for biometric authentication in virtual reality environments. In *2019 ieee international conference on artificial intelligence and virtual reality (aivr)* (p. 9-97). Los Alamitos, CA, USA: IEEE Computer Society. Retrieved from `https://doi.ieeecomputersociety.org/10.1109/AIVR46125.2019.00012`

Ataş Musa. (2017). Hand Tremor Based Biometric Recognition Using Leap Motion Device. *IEEE Access*, *5*, 23320–23326. Retrieved from `http://dx.doi.org/10.1109/ACCESS.2017.2764471`

Bangor, A., Kortum, P., & Miller, J. (2009, May). Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, *4*(3), 114–123.

Bhalla, A., Sluganovic, I., Krawiecka, K., & Martinovic, I. (2021). Movear: Continuous biometric authentication for augmented reality headsets. In *Proceedings of the 7th acm on cyber-physical system security workshop* (p. 41–52). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3457339.3457983`

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. Retrieved from `https://doi.org/10.1023/a:1010933404324`

Brostoff, S., & Sasse, M. A. (2003). "Ten strikes and you're out": Increasing the number of login attempts can improve password usability. In *Workshop on human-computer interaction and security systems at chi 2003.* Retrieved from `https://discovery.ucl.ac.uk/id/eprint/19826/`

Buschek, D., De Luca, A., & Alt, F. (2015). Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (p. 1393–1402). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2702123.2702252`

Chan, A., Halevi, T., & Memon, N. (2015). Leap motion controller for authentication via hand geometry and gestures. In T. Tryfonas & I. Askoxylakis (Eds.), *Human aspects of information security, privacy, and trust* (pp. 13–22). Cham: Springer International Publishing. Retrieved from `https://doi.org/10.1007/978-3-319-20376-8_2`

Chauhan, J., Kwon, Y. D., Hui, P., & Mascolo, C. (2020, dec). Contauth: Continual learning framework for behavioral-based user authentication. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *4*(4). Retrieved from `https://doi.org/10.1145/3432203`

Chen, Y., Yang, Z., Abbou, R., Lopes, P., Zhao, B. Y., & Zheng, H. (2021). User authentication via electrical muscle stimulation. In *Proceedings of the 2021 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3411764.3445441`

Corner, M. D., & Noble, B. D. (2002). Zero-interaction authentication. In *Proceedings of the 8th annual international conference on mobile computing and networking* (p. 1–11). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/570645.570647`

Deutschmann, I., Nordström, P., & Nilsson, L. (2013). Continuous authentication using behavioral biometrics. *IT Professional*, *15*(4), 12-15.

Dick, E. (2021). *Balancing user privacy and innovation in augmented and virtual reality* (Tech. Rep.). Information Technology and Innovation Foundation.

Gamboa, H., & Fred, A. (2004). A behavioral biometric system based on human-computer interaction. In A. K. Jain & N. K. Ratha (Eds.), *Biometric technology for human identification* (Vol. 5404, pp. 381 – 392). SPIE. Retrieved from `https://doi.org/10.1117/12.542625`

George, C., Khamis, M., von Zezschwitz, E., Schmidt, H., Burger, M., Alt, F., & Hußmann, H. (2017). Seamless and secure VR: Adapting and evaluating established authentication systems for virtual reality. In *Proceedings 2017 workshop on usable security.* San Diego, CA, USA: Internet Society. Retrieved from `https://doi.org/10.14722/usec.2017.23028`

Groenewald, C., Anslow, C., Islam, J., Rooney, C., Passmore, P. J., & Wong, B. (2016). Understanding 3d mid-air hand gestures with interactive surfaces and displays: a systematic literature review. *Electronic Workshops in Computing (eWiC).*

Hamid, L. (2015). Biometric technology: not a password replacement, but a complement. *Biometric Technology Today*, *2015*(6), 7–10.

Han, S., Liu, B., Cabezas, R., Twigg, C. D., Zhang, P., Petkau, J., ... Wang, R. (2020, July). Megatrack: Monochrome egocentric articulated hand-tracking for virtual reality. *ACM Trans. Graph.*, *39*(4). Retrieved from `https://doi.org/10.1145/3386569.3392452`

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904-908. Retrieved from `https://doi.org/10.1177/154193120605000909`

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in psychology* (pp. 139–183). Amsterdam, The Netherlands: Elsevier. Retrieved from `https://doi.org/10.1016/s0166-4115(08)62386-9`

Jain, A. K., Flynn, P., & Ross, A. A. (2007). *Handbook of biometrics.* Springer Science & Business Media.

Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). *Introduction to biometrics.* Springer Science & Business Media.

Jakobsson, M., Shi, E., Golle, P., & Chow, R. (2009). Implicit authentication for mobile devices. In *Proceedings of the 4th usenix conference on hot topics in security* (p. 9). USA: USENIX Association.

John Brooke. (1996). Sus: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability evaluation in industry* (pp. 189–195). London: CRC Press. Retrieved from `https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke`

Kamaishi, S., & Uda, R. (2016). Biometric authentication by handwriting using leap motion. In *Proceedings of the 10th international conference on ubiquitous information management and communication.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2857546.2857583`

Kupin, A., Moeller, B., Jiang, Y., Banerjee, N. K., & Banerjee, S. (2019). Task-driven biometric authentication of users in virtual reality (vr) environments. In I. Kompatsiaris, B. Huet, V. Mezaris, C. Gurrin, W.-H. Cheng, & S. Vrochidis (Eds.), *Multimedia modeling* (pp. 55–67). Cham: Springer International Publishing.

Li, S., Ashok, A., Zhang, Y., Xu, C., Lindqvist, J., & Gruteser, M. (2016, March). Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns. In *2016 ieee international conference on pervasive computing and communications (percom)* (p. 1-9).

Liebers, J., Abdelaziz, M., Mecke, L., Saad, A., Auda, J., Gruenefeld, U., ... Schneegass, S. (2021). 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.* New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/3411764.3445528`

Liebers, J., Horn, P., Burschik, C., Gruenefeld, U., & Schneegass, S. (2021). Using Gaze

Behavior and Head Orientation for Implicit Identification in Virtual Reality. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (pp. 1–9). New York, NY, USA: Association for Computing Machinery.

Liebers, J., & Schneegass, S. (2020). Introducing functional biometrics: Using body-reflections as a novel class of biometric authentication systems. In *Extended abstracts of the 2020 chi conference on human factors in computing systems* (p. 1–7). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3334480.3383059

Manabe, T., & Yamana, H. (2019). Two-factor authentication using leap motion and numeric keypad. In A. Moallem (Ed.), *Hci for cybersecurity, privacy and trust* (pp. 38–51). Cham: Springer International Publishing.

Maruyama, K., Shin, J., Kim, C. M., & Chen, C.-L. (2017). User authentication using leap motion. In *Proceedings of the international conference on research in adaptive and convergent systems* (p. 213–216). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3129676.3129698

Miller, M. R., Herrera, F., Jun, H., Landay, J. A., & Bailenson, J. N. (2020, October). Personal identifiability of user tracking data during observation of 360-degree VR video. *Scientific Reports*, *10*(1). Retrieved from https://doi.org/10.1038/s41598-020-74486-y

Miller, R., Ajit, A., Kholgade Banerjee, N., & Banerjee, S. (2019, Dec). Realtime behavior-based continual authentication of users in virtual reality environments. In *2019 ieee international conference on artificial intelligence and virtual reality (aivr)* (p. 253-2531). San Diego, CA, USA: IEEE.

Miller, R., Banerjee, N. K., & Banerjee, S. (2020). Within-system and cross-system behavior-based biometric authentication in virtual reality. In *2020 ieee conference on virtual reality and 3d user interfaces abstracts and workshops (vrw)* (p. 311-316). Atlanta, Georgia, USA: IEEE.

Miller, R., Banerjee, N. K., & Banerjee, S. (2022). Temporal Effects in Motion Behavior for Virtual Reality (VR) Biometrics. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 563–572). IEEE.

Monrose, F., & Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, *16*(4), 351-359. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167739X9900059X

Mustafa, T., Matovu, R., Serwadda, A., & Muirhead, N. (2018). Unsure how to authenticate on your vr headset? come on, use your head! In *Proceedings of the fourth acm international workshop on security and privacy analytics* (p. 23–30). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3180445.3180450

O'Gorman, L. (2003). Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, *91*(12), 2021–2040.

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

Olade, I., Liang, H.-N., Fleming, C., & Champion, C. (2020). Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (vr). In *Proceedings of the 2020 4th international conference on virtual and augmented reality simulations* (p. 45–52). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3385378.3385385

Pearlman, K. (2020, January). Virtual reality brings real risks: Are we ready? San Francisco, CA: USENIX Association.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Édouard Duchesnay (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. Retrieved from http://jmlr.org/papers/v12/pedregosa11a.html

Pfeuffer, K., Geiger, M. J., Prange, S., Mecke, L., Buschek, D., & Alt, F. (2019). Behavioural biometrics in vr: Identifying people from body motion and relations in virtual reality. In *Proceedings of the 2019 chi conference on human factors in computing systems* (p. 1–12).

New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3290605.3300340

Saevanee, H., & Bhatarakosol, P. (2008, Dec). User authentication using combination of behavioral biometrics over the touchpad acting like touch screen of mobile device. In *2008 international conference on computer and electrical engineering* (p. 82-86).

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

Schneegass, S., Oualil, Y., & Bulling, A. (2016). Skullconduct: Biometric user identification on eyewear computers using bone conduction through the skull. In *Proceedings of the 2016 chi conference on human factors in computing systems* (p. 1379–1384). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2858036.2858152

Shen, Y., Wen, H., Luo, C., Xu, W., Zhang, T., Hu, W., & Rus, D. (2019, May). Gaitlock: Protect virtual and augmented reality headsets using gait. *IEEE Transactions on Dependable and Secure Computing*, *16*(3), 484-497.

Sivasamy, M., Sastry, V. N., & Gopalan, N. P. (2020). Vrcauth: Continuous authentication of users in virtual reality environment using head-movement. In *2020 5th international conference on communication and electronics systems (icces)* (p. 518-523). Coimbatore, India: IEEE.

Sugrim, S., Liu, C., & Lindqvist, J. (2019, September). Recruit until it fails: Exploring performance limits for identification systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *3*(3). Retrieved from https://doi.org/10.1145/3351262

Traoré, I., & Ahmed, A. A. E. (2012). Introduction to Continuous Authentication. In I. Traore & A. A. E. Ahmed (Eds.), *Continuous Authentication Using Biometrics: Data, Models, and Metrics* (pp. 1–22). Hershey, PA, USA: IGI Global. Retrieved from https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-61350-129-0.ch001

Wang, X., & Zhang, Y. (2021). Nod to auth: Fluent ar/vr authentication with user head-neck modeling. In *Extended abstracts of the 2021 chi conference on human factors in computing systems.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3411763.3451769

Weichert, F., Bachmann, D., Rudak, B., & Fisseler, D. (2013). Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, *13*(5), 6380–6393. Retrieved from https://www.mdpi.com/1424-8220/13/5/6380

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 143–146). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1978942.1978963

Xiao, G., Milanova, M., & Xie, M. (2016). Secure behavioral biometric authentication with leap motion. In *2016 4th international symposium on digital forensic and security (isdfs)* (p. 112-118).

Yu, Z., Liang, H., Fleming, C., & Man, K. L. (2016). An exploration of usable authentication mechanisms for virtual reality systems. In *2016 ieee asia pacific conference on circuits and systems (apccas)* (p. 458-460). Jeju, South Korea: IEEE.