Building and Evaluating Anonymizations for Human Motions

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

von der KIT-Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Simon Hanisch

Tag der mündlichen Prüfung: 19.01.2026

1. Referent:	Prof. Dr. Thorsten Strufe
2. Referent:	Prof. Dr. Marc Langheinrich

Karlsruher Institut für Technologie
Fakultät für Informatik
Postfach 6980
76128 Karlsruhe

Abstract

Human body motion is a rich source of information. Capturing motion data opens up a range of many new applications in mixed reality, robotics, and medicine. With the proliferation of inexpensive motion tracking hardware, such as mixed reality headsets, motion tracking suits, and smartphones, this source of information is becoming increasingly accessible in our everyday lives. Combined, this data can be used to create an accurate motion profile of the person captured, allowing for the creation of digital twins, the motion control of robots, and for more accurate diagnoses in medicine. However, motion data is also an inherently sensitive source of information, as it is behavioral biometric data that allows for many privacy sensitive inferences about the captured individual. An attacker with access to someone's motion data could identify that person by their unique motion patterns, infer the presence of diseases, such as Parkinson's, or infer private attributes such, as sex and weight.

In this thesis, we seek to understand the privacy problems of motion data and how we can solve them to enable the privacy preserving usage of motion data.

Since privacy for motion data is a new research topic, we began our investigation with an extensive literature review of privacy-preserving techniques for behavioral biometric data.

The main findings of the literature review are that there are only preliminary approaches to anonymizing motion data, that the methodology for evaluating biometric data anonymization is limited and needs improvement, and that there are few datasets suitable for evaluating the efficiency of motion data anonymization.

Next, we investigated motion anonymization using various simpler techniques and their combinations to better understand which techniques would be effective for anonymizing motion data. Using gait data, we found that identifying individuals is very resilient and difficult to prevent, even when data utility is severely degraded. Due to the high number of correlations in motion data, anonymizing it is a hard problem. We also investigated the privacy issues surrounding motion data by collecting FacialMotionID, the first comprehensive dataset of facial motions. Using this dataset, we demonstrated that facial motion data can also be used to identify individuals.

We addressed the lack of a strong evaluation methodology for biometric data anonymization. We developed a stronger attacker model based on stronger assumptions about the attacker's capabilities. This makes our evaluation methodology more rigorous and produces more reliable privacy results.

In order to address the lack of motion datasets, we collected the gait sequences of 50 people using an IMU motion capture suit to create CeTI-Locomotion.

Lastly, we present Pantomime, the first general anonymization technique for full-body motion data that can maintain high utility while removing most identifiable information. Pantomime demonstrates that meaningful anonymization of motion data is possible, while keeping the utility of the data high.

Acknowledgments

Throughout my thesis I was supported by many great people who helped me finishing this project.

First, I like to thank Thorsten Strufe, my supervisor and mentor. He is the reason I pursued a PhD in the first place. I would like to thank him for providing a great work environment for my thesis, freeing up my time by handling grant proposals, and letting me pursue my own ideas and research directions. He has always given me valuable feedback on my work, helping me become an effective researcher. Thank you for convincing me to pursue a PhD in the first place.

I like to thank my second reviewer Prof. Dr. Marc Langheinrich for accepting to review my dissertation and taking the time to grade my work.

Then I like to thank all of my collaborators with whom I had the pleasure to write papers together during my PhD time. Amr Osman, who was my supervisor during my master thesis and with whom I then wrote my first academic paper publishing the results of the thesis. Patricia Arias-Cabarcos and Javier Parra-Arnau for teaching me how to write a proper survey and always giving me good feedback. During my PhD, I had the pleasure of working with neuroscientists on interdisciplinary research projects. I would like to thank Evelyn Muschter, who co-authored my first biometric recognition paper with me and helped me navigate the neuroscience aspects of identifying people from their motion data, as well as teaching me how different fields write their papers. I would also like to thank Loreen Pogrzeba, from whom I learned a great deal about designing and executing proper data collection studies, as well as collaborating on our social touch demonstrator. Finally, I would like to thank Annika Dix, with whom I had the pleasure of building virtual reality hotwire experiments. Furthermore, I would like to thank Daniel Shea, the textician who helped me refine my methodology paper. A big thank you goes to Julian Todt for the many papers we wrote together, and for supporting me on my ’short’ detour to try and improve the state of the art in evaluation methodology for biometric data anonymization.

I would like to thank my fellow PhD students — Felix, Àlex, Patricia, Daniel, Marcel, Fritz, Julian, Kamyar, Matin, Shima, Christoph, Christiane and Jan — who made coming to the office always fun and who supported me through the highs and lows of PhD life.

Furthermore, I would like to thank Mrs. Sauer, our secretary, and Marius Simianer, our technician, for their invaluable assistance in managing systems and navigating university bureaucracy.

I would like to thank my parents for all their support throughout my education journey from school to PhD and for not asking me too often how long my PhD is going to take.

Finally, I would like to thank my girlfriend, Josephine, whom I love dearly, for her support throughout my PhD journey, her patience with the late nights and her courage in following me to Karlsruhe.

Contents

List of Figures

List of Tables

1. Introduction

Human motion data is a rich source of information with applications in many different fields, including medicine, robotics, and mixed reality. In medicine, for example, motion data has been used for the diagnosis of diseases like Parkinson's from gait data [2] or for the monitoring of therapy progress [15, 315]. In robotics, human motion data is important because it can be used for transfer learning, in which humans perform motions that are mimicked by robotic agents [64, 349]. Using human examples speeds up the learning process and makes it is easier to generate training data to teach the robots. Lastly, in the field of mixed reality, human motion data is an indispensable means to control the devices (e.g. hand motion control [232, 17]) and to enrich its applications (e.g. mixed reality chats with body motions [368]). These mixed reality applications have even spawned a new streaming subgenre: virtual YouTubers (VTubers)[1]. VTubers use motion tracking to animate digital avatars (often anime characters) that they act out online. This trend is ongoing and, for the first time, surpassed 500 million hours watched in the first quarter of 2025[2].

In the past, motion tracking was complicated and expensive, limiting its use to expert applications, like movie production and medical labs. Now, new technologies like inside-out tracking using RGB cameras and the usage of *Inertial-Measurement-Unit* (IMU)-sensors for full body tracking [312, 241], combine reliability, a low price tag (below 300€), and ease of use. These advances have led to a fast proliferation of motion tracking and puts motion tracking capabilities into the hands of consumers around the world. Furthermore, today's machine learning-driven extraction capabilities can be used to extract 3D motion data from

[1] https://en.wikipedia.org/wiki/VTuber

[2] https://streamscharts.com/news/vtubers-q1-2025-report

This chapter is based on the contributions:

- **Simon Hanisch**, Patricia Arias-Cabarcos, Javier Parra-Arnau, and Thorsten Strufe. "Anonymization Techniques for Behavioral Biometric Data: A Survey". In: ACM Computing Surveys. 2025. DOI: 10.1145/3729418.
- **Simon Hanisch**, Evelyn Muschter, Admantini Hatzipanayioti, Shu-Chen Li, and Thorsten Strufe. "Understanding Person Identification Through Gait". In: Proceedings on Privacy Enhancing Technologies. 2023. DOI: 10.56553/popets-2023-0011.
- **Simon Hanisch**, Loreen Pogrzeba, Evelyn Muschter, Shu-Chen Li, and Thorsten Strufe. "A kinematic dataset of locomotion with gait and sit-to-stand movements of young adults". In: Scientific Data 11.1 2024. DOI: 10.1038/s41597-024-04020-6.
- **Simon Hanisch**, Julian Todt, Jose Patino, Nicholas W. D. Evans, and Thorsten Strufe. "A False Sense of Privacy: Towards a Reliable Evaluation Methodology for the Anonymization of Biometric Data". In: Proceedings on Privacy Enhancing Technologies. 2024. DOI: 10.56553/popets-2024-0008.
- **Simon Hanisch**, Julian Todt, and Thorsten Strufe. "Pantomime: Towards the Anonymization of Motion Data using Foundation Motion Models". 2025. DOI: 10.48550/arXiv.2501.07149

simple RGB videos [306]. This means that any type of video can be used as a source of motion data.

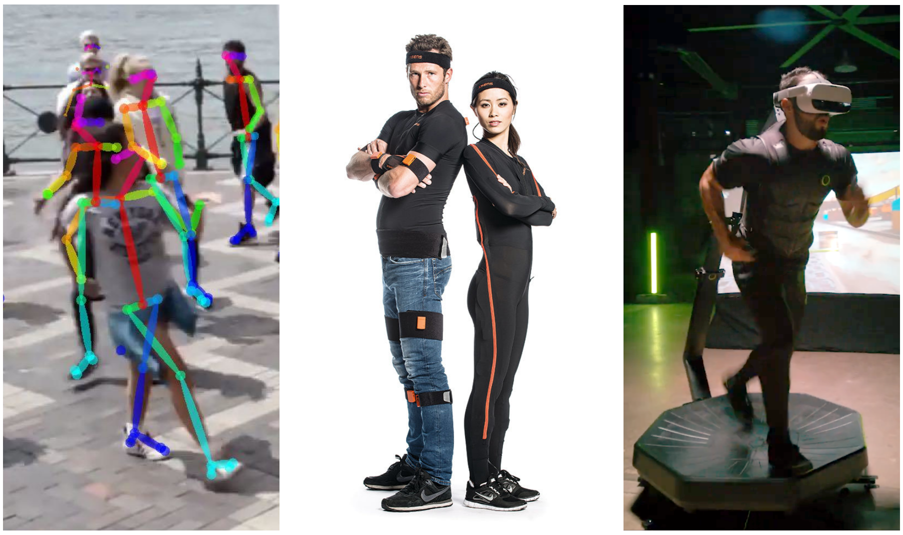


Figure 1.1.: Variety of recent motion capturing devices recording gait of individuals in public and private spaces from video or inertial measurement units (cf. OpenPose, VFXVoice, Virtuix).

With motion data leaving specialized fields, we must increasingly consider its disadvantages, as it poses an inherent privacy risk to people captured. Motion data is behavioral biometric data that can be used to identify individuals based on how they walk [115, 134, 369] (gait), how they move their eyes [173], or how they move their head and arms [250, 240]. Using gait data, individuals can be identified from both low resolution 2D RGB videos [369], as well as 3D motion captures [115], showcasing that gait recognition poses a similar threat to privacy as face recognition. Besides identification, motion data can be used to infer private attributes about individuals. For example, the motion data of individuals can be used to infer their age [365], sex [365], or medical status [186].

These privacy problems are all the more pressing when we consider how motion data will be used. Take, for example, the field of mixed reality, where motion data is used for the animation of digital avatars. When these avatars are used on platforms like the Metaverse[3] motion data is shared both with the platform, app providers and all the people who can see the animated avatar. Here, the motion data is implicitly published and malicious actors can gain access to it.

In order to address these privacy problems anonymization techniques are required when motion data is shared with services or other people. An anonymization technique is a privacy enhancing technique that modifies original data to reduce the risk of privacy inferences, such as identification and attribute inferences, while retaining the data's usefulness for its intended purpose. Hence, an anonymization always seeks to achieve two goals, privacy and utility. As these goals are often contrary to each other, a privacy-utility trade-off must be made during the design of an anonymization.

[3] https://en.wikipedia.org/wiki/Metaverse

In this thesis we investigate the two problems of building and evaluating anonymizations for motion data. First, we survey the literature of existing behavioral biometric data anonymizations to understand which solutions already exist and to see which techniques from other behavioral biometric traits can be adapted for motion data. For this we follow the methodology of Kitchenham at al. [166] to perform a systematic literature survey for the behavioral biometric traits of gait, hand motions, eye gaze, voice, brain activity, and heartbeat. To allow for a better comparability between behavioral biometric traits, we propose a unified taxonomy for behavioral biometric anonymizations based on how they modify the original data to ensure privacy.

We then study how robust the identifying information in gait data is, and how difficult it is to remove. To this end, we apply different simple anonymization techniques, as well as combinations thereof, to features that could potentially allow identification in the data. Our goal is to anonymize motion data by focusing only on features that are important for identification. We conducted further research into identifying individuals from abstract facial motion data. To this end, we collected the first facial motion dataset using *Mixed Reality* (MR) headsets.

We further improved the evaluation methodology for anonymization techniques for biometric data and, therefore, for motion data as well. Our evaluation methodology focuses on creating a strong adversary that seeks to identify individuals in anonymized data. Different to prior work, the adversary is aware of the anonymization in place and adapts their attack to it. Additionally, the attacker reduces the number of individuals and selects the ones that are the most difficult to anonymize, creating a more challenging evaluation scenario for the anonymization under test.

After reviewing the literature on identifying individuals through gait and determining which features are important, we focus on collecting motion data from different sources, as this is necessary for developing anonymization techniques. We conducted a gait study in which the motions of 50 people in Karlsruhe (+30 people in Dresden) were recorded using IMU-suits. The participants had to perform four different gait task (e.g., carrying a crate, walking fast) and one sit-to-stand task with a high number of repetitions.

Lastly, we propose Pantomime, an approach for anonymizing body motion data using foundation motion models. Pantomime first transforms the body motion data from various sources into a normalized format. Then, it anonymizes the data in the latent space of a foundation motion model by adding noise to the original latent code. Due to the anonymization occurring in the latent space of a foundation model, the resulting anonymized latent code decodes back to a valid human motion.

Our contributions in this thesis are as follows:

- We perform a survey of the current anonymization techniques for behavioral biometric data and propose our own taxonomy to classify the anonymization techniques based on how they modify the original data across behavioral biometric traits.

- We investigate which features enable the identification of individuals via gait recognition.

- We were the first to investigate the use of abstract facial motion data for identification purposes.

- We improve the evaluation methodology for biometric data anonymization by creating a more challenging anonymization scenario.

- We collect a gait dataset to evaluate behavioral biometric data anonymizations.
- We propose Pantomime, a full-body motion anonymization that uses foundation motion models to anonymize motion data.

The thesis is structured as follows: Chapter 2 introduce the background and Chapter 3 presents our survey. In Chapter 4 we investigate which features contribute to the identification of individuals via gait. In Chapter 5 we present a novel study on identifying people via abstract facial motions. We then describe our improvements to the evaluation methodology in Chapter 6. Next, we present our motion data collection in Chapter 7. Lastly, we present Pantomime in Chapter 8, before drawing our conclusion in Chapter 9.

1.1. Collaborations

During my thesis work, I had the opportunity to collaborate with many co-authors on various papers. In this work, I will use the academic "we" to honor these collaborations because I would not have been able to conduct my research without them. Here, I clarify which collaborations contributed to which parts of the papers on which this thesis is based.

Thorsten Strufe, my supervisor, collaborated with me on all but one paper("Side-Channel Attacks on Query-Based Data Anonymization”) and helped me with discussions of my research, gave countless feedback on my ideas, and editing and writing advice for my papers.

Julian Todt was a student of mine and is now a valued colleague. We worked closely together to improve the evaluation methodology for biometric data anonymization. He ran parallel experiments on face recognition approaches for our methodology papers and helped develop the selection strategies, which were a key part of our work. Julian also assisted me with writing and editing many of my papers.

Evelyn Muschter is a co-author of my first paper on biometric data anonymization. Coming from the field of neuroscience, she helped me understand how humans identify people by their motion cues. She also helped with my data collection study, in which we recorded the gaits of 50 people performing various exercises. Loreen Pogrzeba also helped prepare the user study, process the data, and write the paper.

Both Patricia Arias-Cabarcos and Javier Parra-Arnau contributed to the survey on the anonymization of behavioral biometric data. They assisted in writing the introduction, methodology, heartbeat, and brain activity sections. They also helped edit and improve the rest of the survey.

Adriano Castro did a bachelor’s thesis with me, which focused on identifying people based on their facial motions. Together, we designed the study, and he implemented the apparatus for data collection and helped with post-processing the data.

2. Background

In this chapter, we provide an overview of the terminology used throughout the thesis. We also introduce the most important metrics and machine learning concepts that we will use in this thesis. Lastly, we describe the human gait cycle and introduce point-light displays.

2.1. Terminology

Biometric traits (also called biometric characteristics [143]) are properties of a human that either capture the physiology of a human (e.g. face, iris, fingerprint) or their behavior (e.g. voice, gait, heartbeat). Human motion, while also containing some physiology information like height, is primarily a behavioral biometric trait, that is is recorded as a time series.

Due to the unique nature of biometric traits for each human being they can be used for privacy-invasive inferences. We distinguish between two privacy threats. By the term **identity inference**, we mean that the identity of an individual is inferred. By the term **attribute inference**, we mean that only a specific private attribute (e.g. age, sex, medical condition) is inferred.

In biometric recognition, identity inference and attribute inference are made operative in a system that learns an inference on representative samples for each class. For each biometric sample to be classified, the biometric recognition system returns a list of possible classes, where each class has been assigned its own separate likelihood. In closed-set recognition, the sample must belong to one of the classes in the dataset, while in open-set recognition the sample may belong to an unknown class.

To prevent biometric recognition **privacy enhancing technologies (PETs)** are employed which obfuscate the private information in the data from internal and external observers. The specific term of **anonymization** refers to PETs which aims to remove all identifiers that directly identify individuals. Anonymization takes biometric **clear data** as input and outputs **anonymized data**.

The aim of anonymization is to protect an individual's identity. During the process of anonymization, information is removed or perturbed that is specific to an individual. Hence, anonymization prevents an adversary from using the data to infer the class corresponding to an individual (i.e. identification). In contrast to anonymization, **pseudonymization** is

This chapter is based on the contributions:

- **Simon Hanisch**, Patricia Arias-Cabarcos, Javier Parra-Arnau, and Thorsten Strufe. "Anonymization Techniques for Behavioral Biometric Data: A Survey". In: ACM Computing Surveys. 2025. DOI: 10.1145/3729418.
- **Simon Hanisch**, Julian Todt, and Thorsten Strufe. "Pantomime: Towards the Anonymization of Motion Data using Foundation Motion Models". 2025. DOI: 10.48550/arXiv.2501.07149

aimed at retaining some connection between identity and data in order to link the data to an alternative identifier.

We shall employ the term **utility** to quantify the degree of functionality maintained concerning a service for which the behavioral biometric data is intended. The utility is kept despite the implementation of a PET that may hide or perturb part of the data which may degrade the quality of the service. We stress that utility in this context does not refer to user-interface design. Depending on each application, the behavioral biometric data is utilized for a variety of purposes. For example, in an application for biometric authentication, an evident measure of utility is the ability to verify the identity of an individual. Likewise, in an application based on human computer-interaction, we may require the behavior to still work as reliable input modality for computer systems. In a healthcare, application we may be interested in detecting abnormal behavior patterns, and monitoring specific aspects of the behavior such as counting steps or inferring the preferences of a user for personalization. The utility of the provided service may be assessed as the performance in carrying out those tasks.

As pointed out above in the introduction, any PET poses a **trade-off between privacy and utility**. The optimization of the privacy-functionality (or privacy-utility) trade-off will refer to the design and tuning of PETs in order to maximize privacy for a desired functionality, or vice versa.

2.2. Machine Learning

Since we are investigating the anonymization of motion data, we often rely on biometric recognition systems to evaluate the privacy of the anonymizations. These systems, in turn, are based on machine learning algorithms that perform a classification task by separating biometric samples into user classes. For Pantomime, our proposed motion data anonymization, we use foundation models with an autoencoder structure. Below, we provide a brief overview of the necessary concepts.

2.2.1. Metrics

Here, we define the metrics used to measure the performance of classification evaluations of privacy using biometric recognition systems. These metrics are defined in terms of the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy is defined as:

$$accuracy = \frac{TP + TN}{FN + FP + TP + TN}$$

Due to imbalanced classes in some datasets, we also use balanced accuracy, which treats every class equally. It is defined as follows:

$$balanced\ accuracy = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right)$$

As an alternative we also use the F_1 score, which is defined as follows:

$$F_1 = \frac{2TP}{2TP + FP + TN}$$

For authentication evaluations, the *Equal Error Rate* (EER) is often used to understand how well a system performs when the false rejection rate and false acceptance rate are equal. It is defined as follows:

$$EER = \frac{1}{2}\left(\frac{FP}{TP+FP} + \frac{FN}{TN+FN}\right) \quad at \quad \min_t \left|\frac{FP}{TP+FP} - \frac{FN}{TN+FN}\right|$$

A common loss function to use for training *Autoencoder* (AE) is *Mean Squared Error* (MSE), which is defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Given the number of samples N, the original sample y_i and the output sample $\hat{y}_i$.

2.2.2. Principal Component Analysis

A *principal component analysis* (PCA) is a linear dimension reduction technique that represents data in a lower-dimensional space that captures the greatest variance. PCA is based on a covariance matrix, which captures the covariance between each pair of data points. In simple terms, the covariance matrix captures the total linear variance of the dataset. To perform PCA, the covariance matrix is decomposed using an eigenvector analysis to obtain the matrix's principal components. The first vector captures most of the variance in the data. The second vector captures the main variance after removing the first vector, and so on. The original data is represented as a linear combination of these components, which reduces the dimensionality of the data to the number of components used in PCA. Due to its focus on variance between data points and reduced dimensionality, PCA is a common preprocessing step for classification tasks.

2.2.3. Support Vector Machines

A *Support Vector Machine* (SVM) is a supervised machine learning algorithm that is primarily used for classification tasks. It classifies points by fitting a hyperplane to separate points of two different classes. An SVM attempts to identify the maximum distance between the hyperplane and the support vectors, which are the points closest to the hyperplane. To separate data points that cannot be separated linearly, an SVM uses kernels. These are functions that map the data points to a higher-dimensional space in which the points can be separated. This process is known as the kernel trick. In this thesis, we use a radial basis function (RBF) kernel, which computes the squared Euclidean distance between two points, multiplies the result by a free parameter $-\gamma$, and then computes the exponential function of the result. Since SVMs can only separate two classes of points, $n - 1$ binary one-vs-all classifiers must be trained to apply them to multi-class classification problems.

2.2.4. Autoencoders

For our proposed motion data anonymization method, Pantomime, we rely on foundation models that use variants of AE architectures and some of their features. This is because the anonymization process takes place in latent space. Here, we focus on the important features of AE and its variants for Pantomime.

An *Autoencoder* (AE) [171] is a machine learning architecture consisting of an encoder and a decoder. The encoder translates the data into a much smaller latent space, and the decoder translates it back into the original data space. In other words, the encoder compresses the data, and the decoder decompresses it. The model's overall goal is to learn an efficient encoding of the input data. Because the model's output should match the input, it can be trained unsupervised using a reconstruction loss (e.g., MSE) between the encoder input and the decoder output.

A *Variational Autoencoder* (VAE) [164] is a type of AE that instead of learning a discrete latent code for a given input, maps the input to the parameters of a probability distribution. From this probability distribution a discrete latent code is then drawn and decoded to the input space by the decoder. Often when training a VAE the shape of the learned latent distribution is regularized by using a Kullback-Leibler divergence [175] to be similar to a normal distribution. In this case the learned latent space can be interpreted as a mixture of normal distributions.

There exist different variants of VAEs. β-VAEs [124] use a weighted Kullback-Leibler divergence to increase its influence in the loss and by that force a disentanglement of the dimensions of the latent code [337, 43]. Another type of VAE is a *Conditional Variational Autoencoder* (CVAE) [337], which uses an additional label to constrain the latent code of an input to be deterministic. In this way, only the label can be fed to the decoder to generate a sample that belongs to the class of the input label.

3. Literature Survey

The ongoing digital transformation is leading to an increasingly comprehensive data collection on citizens. Ever improving peripherals, like augmented reality (AR)/virtual reality (VR) goggles, motion capturing suits and gloves, force-feedback input devices, sensor-rich cell phones, smart watches, and other wearables drastically increase the coverage and resolution at which biometrics and behavioral data of individuals become available for processing.

Preserving the privacy, and ultimately the dignity of individuals who come in the range of sensors and are captured in their behavior requires more sophisticated approaches than removing direct identifiers (IP address, social security number (SSN), blurring a face) or intuitive quasi identifiers (gender, age, ethnicity) in databases.

Anonymizations techniques for biometric traits have long been a topic of research with a large focus on anonymization of physical biometric traits such as faces, or irises. Unsurprisingly, today a large corpus of literature exist on the topic with multiple literature surveys [311, 227, 331] cataloging and summarizing the current state-of-the-art. However, human motion data is not a physical biometric trait but a behavioral one, differing the important fact that instead of only being captured at a single point in the time but as a time-series which captures change. While some anonymization techniques, such as blurring or pixelation, can be adapted to be used for behavioral biometric traits often new techniques are required which accommodate to the specific requirements of behavioral biometric traits, such as the consistency across every timestep. Hence, new anonymization techniques have been developed for behavioral biometric traits.

A growing corpus of studies is addressing this challenge of anonymizing behavioral data. They focus on a variety of different human traits, ranging from the voice, over gait, to less prominent examples like gestures, heartbeat, and others. A systematic review of all these approaches, which bridges the attempts in extracting the shared conceptual and methodological similarities is missing, to the best of our knowledge. Further, we want to highlight both differences between approaches, their conceptual properties, as well as future research opportunities.

For this chapter, we hence set out to systematize the corresponding literature. We are interested in anonymizations for scenarios in which behavioral data is collected by or shared with third parties to perform a specific operation. As we are interested more in privacy than confidentiality we do not consider approaches in which an entity encrypts its own data to hide it from access by unintended audiences. We are rather interested in approaches that protect from unintended revelation of information contained in data [61].

This chapter is based on the contribution:

- **Simon Hanisch**, Patricia Arias-Cabarcos, Javier Parra-Arnau, and Thorsten Strufe. "Anonymization Techniques for Behavioral Biometric Data: A Survey". In: ACM Computing Surveys. 2025. DOI: 10.1145/3729418

We deem 'confidential computing', processing based on homomorphic cryptography, or similar approaches in which the data owner is the only entity that learns anything from the data, out of scope of our analysis. For our study, we followed Kitchenham's guidelines [166] to systematically discover and survey the current state of the art, comprising of 142 distinct studies, extracted from a corpus of 364 initially discovered publications.

We identify common applications that process behavioral data, to extract sensible measures of utility, as well as common privacy threats with corresponding adversary models. We define two taxonomies of anonymization approaches. The first is defined by how the anonymization transforms the data and the second by which anonymization goal it seeks to protect. Next, we provide a detailed overview of the different anonymization approaches, sorted by the trait they aim to protect. We provide insight into the corresponding applications that define the utility, and into the privacy threats, privacy goals, applied anonymization concepts, and the evaluation the corresponding scientists performed, together with the data they chose for their studies.

As a main findings we show how the underlying anonymization concepts are independent of the biometric trait. In consequence, we identify biometric traits for which specific anonymization concepts have not yet been tested. Further, we find that the general evaluation methodology for behavioral biometric anonymization implies a weak adversary and must be improved to convincingly assess the efficacy of protective measures.

The main contributions of this work are as follows:

- Following Kitchenham's guidelines [166] we systematically discovered a corpus 364, which we filtered to 142 distinct proposals for the privacy protection of behavioral biometrics.
- We categorized the works by using two novel taxonomies, allowing the comparison of anonymizations across biometric traits.
- Further, we find that the underlying privacy protection concepts and the general evaluation methodology for behavioral biometrics are independent of biometric traits. This allows novel behavioral biometric traits, such as human motion data, to adapt concepts and methodologies from more established ones like voice.

3.1. Scenario

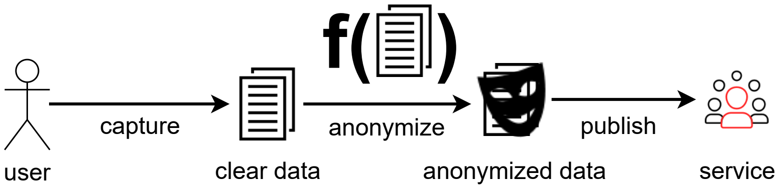


Figure 3.1.: The data-publishing scenario of this thesis.

In this thesis, we assume a data publishing scenario (see Figure 3.1) in which data is first anonymized, then published or processed by a service or application, or shared with one. This includes involuntary publication, which can occur when, for example, the biometric templates of an authentication system are leaked or fitness tracker data is sold.

We assume that the utility of the modified, protected data is preserved, so that the received service (e.g., a personalized recommendation or a played virtual reality game) remains meaningful.

Examples of this scenario include Metaverse applications, in which users represent themselves in digital worlds with animated avatars. In these applications, users' motion data is captured and streamed to the platform (the Metaverse service provider) as well as to other users in the same digital space. Therefore, the motion data is implicitly published to the service provider and third-party users.

In general, the field of human-computer interaction captures and processes behavioral biometric data because each input over time comprises a behavior. Today, touch gestures, keystroke patterns and mouse movements are the main input modalities for computer systems; however, new input modalities, such as voice, and gestures, are on the rise and will likely become more relevant in the coming years. Mixed reality is important in this regard because it combines many of these input modalities and requires constant monitoring of its users.

Another area in which behavioral data is useful is healthcare and the quantified self. Advances in sensors and machine learning techniques have enabled the development of activity recognition, fall detection, and remote health monitoring applications that facilitate the care of elderly, sick, or disabled individuals and ease diagnosis [267, 286, 186]. Typically, data collected includes gait and motion information from accelerometers and gyroscopes embedded in user devices, as well as biosignals such as heartbeat and brain activity. This data can also be processed to provide users with health-related feedback. For example, it can guide users through relaxation exercises or detect and signal cognitive states, such as stress, so users can address them.

3.2. Methodology

We performed a systematic literature review following Kitchenham's guidelines [166] to identify relevant studies on privacy techniques for behavioral data, as it is depicted in Figure 3.2.

Our guiding research question is **"What techniques are applicable to protect behavioral data privacy?"** From this starting point, the goal is to understand how these techniques work, what is the level of protection provided, and what are the limitations and existing open challenges. To answer these questions, we first explored the literature on biometrics [14, 63, 230, 206, 384, 281, 6, 107] to determine what kind of behavioral traits can be used to identify a person. The complete list of behavioral traits we searched includes: brain activity (also referred to as cognitive biometric), eye gaze, facial expression, gait, gesture, handwriting, haptic, heartbeat, keystrokes, lip, motion, mouse, thermal, touch, and voice. Next, we used this list of traits combined with the keyword **"privacy"** and the semantically similar terms **"anonymization"** and **"de-identification"**, as search strings in the main academic databases for computer science. Based on these search terms, we compiled works with no constraints on publication date, obtaining a set of 364 papers spanning from 2007 to October 2024, after filtering duplicates. During pre-screening, we built a taxonomy of privacy solutions and decided to narrow-down the scope of the survey to anonymization techniques focused on protecting the publication of behavioral data from identity and attribute disclosure attacks. We consider approaches that assume collection, sanitization, and subsequent publishing of data, which must be anonymized but also keep

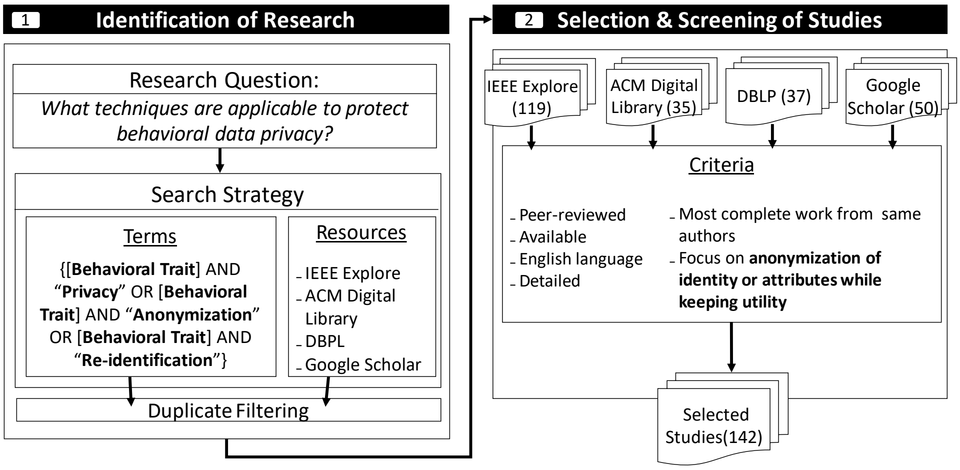

1 Identification of Research
Research Question:
What techniques are applicable to protect behavioral data privacy?
Search Strategy
Terms
{[Behavioral Trait] AND "Privacy" OR [Behavioral Trait] AND "Anonymization" OR [Behavioral Trait] AND "Re-identification"}
Resources
- IEEE Explore
- ACM Digital Library
- DBPL
- Google Scholar
Duplicate Filtering
2 Selection & Screening of Studies
IEEE Explore (119)
ACM Digital Library (35)
DBLP (37)
Google Scholar (50)
Criteria
- Peer-reviewed
- Available
- English language
- Detailed
- Most complete work from same authors
- Focus on anonymization of identity or attributes while keeping utility
Selected Studies(142)

Figure 3.2.: Summary of the procedure for identifying and selecting relevant studies on behavioral data privacy techniques. We first analyzed the literature on biometrics to determine behavioral traits for person identification. We then used these traits as key terms to search for privacy-related publications, following Kitchenham's guidelines for systematic literature reviews [166]. The complete list of behavioral traits we searched includes: brain activity, eye gaze, facial expression, gait, gesture, handwriting, haptic, heartbeat, keystrokes, lip, motion, mouse, thermal, touch, and voice.

a level of utility to provide behavioral data driven services. Accordingly, the down-selection of primary studies to be analyzed in this survey considered the following criteria. Documents were excluded if:

1. The publication format was other than peer-reviewed academic journal or conference paper.
2. The paper could not be retrieved using IEEE Explore, ACM Digital Library, DBLP, or Google Scholar.
3. The publication language was not English.
4. Another paper by the same authors superseded the work, in which case the most complete work was considered.
5. The privacy protection technique was other than identity or attribute anonymization with data utility.
6. The anonymization approach was described at a high level and not enough details were provided to properly address the guiding research question.

The search and selection protocol yielded a final corpus of 142 peer-reviewed works on behavioral data anonymization, which we clustered according to the behavioral trait being protected: gait, brain activity[1], heartbeat, eye gaze, voice, and hand motions (handwriting, keystrokes, mouse movements, and hand gestures). We found no papers on facial expression, lip, touch, and haptic traits that fulfill our criteria.

3.3. Taxonomy

Based on our collected literature corpus and scenario, we identify two main **privacy threats** that apply to behavioral data collected/processed by a third party and can be explained in terms of the related attacker model:

- **Identity Disclosure:** The attacker's goal is to use the behavioral data to identify the user. In this threat, we assume that the attacker is able to link the target's behavioral data to the target's identity and now wants to identify them in another scenario. For example, linking the user account and data in a work-related application to their account in an entertainment application. This linkage would allow the attacker to learn more about the user activity. An example of this type of attacker, as presented in [340], could be a VR headset user entering a federated Metaverse offering several services (e.g., games, adult content, professional training apps). Even if the user tries to use a pseudonym when entering a foreign server, the server and other users can use transmitted behavioral data (e.g., controller/headset motions, eye-tracking) to identify the user across different pseudonyms. Moreover, it is not uncommon that behavioral data is sold to third parties or released unintentionally through a breach or hack[2].
- **Attribute Disclosure:** In this threat, the attacker goal is not to re-identify the user across accounts, but to derive sensitive attributes included within the available behavioral data that the user did not intend to disclose, such as sex, medical conditions, or personal interests. The attacker might have had previous access or could have collected a dataset on which to train the machine learning model for targeted inference. For example, based on publicly available electroencephalogram datasets of alcoholic and non-alcoholic persons [258, 157], it could be possible to build a classifier that determines if newly gathered data from an entertainment application using a brain-computer interface (BCI) belonged to a user with an alcohol problem.

From the privacy threats, we can derive the two **anonymization goals** with which techniques can be categorized, i.e., focused on protecting user **identity** and focused on protecting specific **attributes**, as depicted in Figure 3.3.

- **Identity Protection:** The process of transforming the behavioral biometric data of a person in such a way that their identity can no longer be linked to the data. **Pseudonymization** replaces the identifier of a person with a new one and **anonymization** aims to prevent identification altogether. Given that behavioral biometric data

[1] Brainwave signals are a manifestation of both its physiological structure and the behavioral way it processes information, for example in reaction to stimuli. In the context of this survey, we refer to EEGs as behavioral data, given that this is our main focus of study, but we acknowledge that physiological components are present.

[2] https://www.zdnet.com/article/over-60-million-records-exposed-in-wearable-fitness-tracking-data-breach-via-unsecured-database/

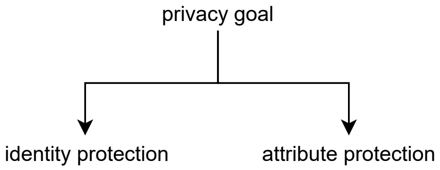


Figure 3.3.: Taxonomy of anonymization techniques for behavioral data protection according to the privacy goal.

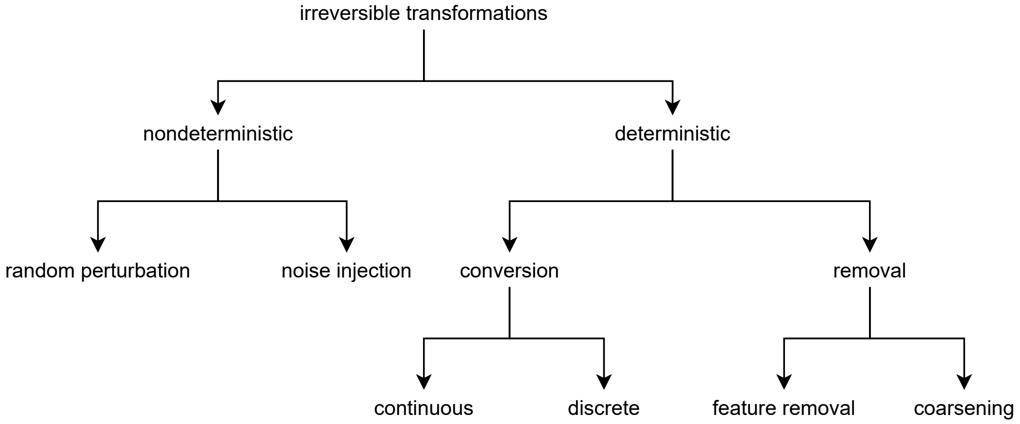


Figure 3.4.: Taxonomy of anonymization techniques for behavioral data protection according to the type of data transformation applied.

is inherently a time-series, the identity of the person is typically maintained throughout the entire time-series. Hence, anonymization of behavioral biometric data usually refers to breaking the link between the identity and a time-series.

- **Attribute Protection:** The process of transforming the behavioral biometric data of a person in such a way that specific private attributes of the person can no longer be inferred from the data. This encompasses both long-living attributes such as age or gender and short-living attributes such as mental state or temporary health conditions. An extreme version of attribute protection is template protection. For **template protection**, the identity verification of the person, in the context of an authentication system, should be still possible while all attributes are protected.

Based on the study of state-of-the-art protection methods, we have conducted a classification of methods that expectedly is not entirely exclusive to the field of behavioral data privacy, as it shares similarities with other classifications in more mature privacy fields, such as statistical disclosure control (SDC). In this section, we elaborate on this classification and establish correspondences with anonymization techniques widely studied in SDC.

Our taxonomy, as depicted in Figure 3.4, of anonymization solutions for behavioral biometric data is based on the **type of transformation** applied to the original data, to derive anonymized, protected data. We include only fundamental concepts, some of the anonymization techniques combine multiple of them. The basic and shared characteristic of all anonymization methods is that they aim to provide irreversible transformations, i.e. it is impossible to transform the data back to the original data.

The first distinction of our taxonomy is if they are deterministic or randomized techniques. **Non-Deterministic methods** rely on randomness in their transformation, which can yield different results for the same input, and **deterministic methods** always give the same result for the identical input. There are several methods under these two approaches, as we detail as follows.

- **Non-Deterministic methods:**
  - **Random perturbations:** A random transformation into a different domain.
  - **Noise injection:** Methods that add random noise to the data points. We find that the corresponding method in the literature of SDC is referred to as *additive noise masking* [141], a perturbative technique that allows for the release of an entire microdata set, where the modified values rather than exact values are released. We would like to emphasize that additive noise masking is combined typically in this field with deterministic transformations.
- **Deterministic methods** are further split into **removal** and **conversion**. The removal method eliminates data points from the data such that the data points do not have an influence on the anonymized result. Conversion methods transform the data points into a new representation, which typically depends on the original domain. The conversion methods are often generalizations of the data.
  - **Removal** can be performed in two ways: **coarsening** and **feature removal**. Coarsening refers to removing parts of each data point or removing detail from the data. Feature removal refers to removing data points belonging to a specific feature altogether. This removal technique is called *suppression* [141] in the SDC field. There, when a microdata set contains too few records sharing a combination of quasi-identifier values, it is termed an "unsafe combination" due to the risk of potential re-identification. To address this concern, specific values of individual variables are deliberately suppressed, and effectively replaced with missing values. This suppression strategy aims to expand the number of records that conform to each combination of key values, thereby eliminating unsafe combinations and enhancing privacy protection.
  - **Conversion** can be **discrete** or **continuous**, depending if the result of the conversion is a discrete or continuous value. As mentioned above in the noise injection technique, SDC also employs transformations of this kind.

3.4. Anonymizations

We have found anonymizations for the behavioral biometric traits of voice, gait, hand motions, eye gaze, brain activity, and heartbeat. Due to our main focus on human motion data we only describe the anonymization techniques for gait and hand motions here in detail. The remaining traits can be found in the Appendix A. For each trait, we look at the utility, threat space, anonymization techniques, and evaluation methodology.

3.4.1. Gait

The human gait is the pattern in which humans move their limbs during locomotion, multiple manners of gait exist such as trotting, walking, or running. Gait can be broken down into individual gait cycles [342] which is the shortest repetitive task during the gait. The gait cycle spans from a specific gait event of one foot until the same foot reaches the same gait event. It consists of a stance phase, in which the foot is on the ground, and a swing phase, in which the foot is in the air. The two phases alternate for each foot.

Due to its usefulness as a behavioral biometric trait for identifying individuals, gait has long been a research interest of both computer science and psychology. For example, Yovel et al. [395] find that it plays an important part for humans to identify people at a distance, and Pollick et al. [288] show that it is possible for humans to infer the gender of a walker, even when the walker is only shown as a set of points, as so-called point-light-display. The following section deals with the anonymization of gait patterns.

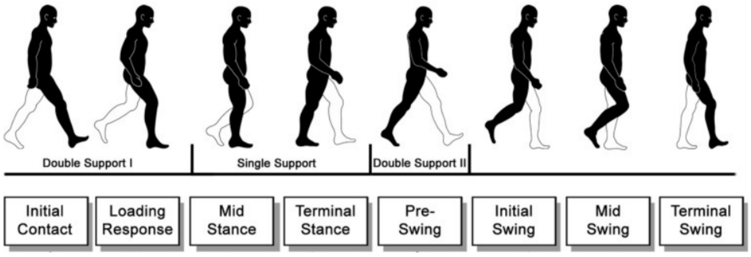


Figure 3.5.: The phases of the gait cycle, source: [342].

Gait recognition methods have been an active research topic in the past, hence a large set of different methods for various capture methods exists. Wan et al. [369] performed a recent survey on the subject and listed recognition methods for cameras, accelerometers, floor sensors, and radars. The main portion of the works focuses on camera based gait recognition which is classified by Wan et al. as either model-based or model-free. Model-based methods use a specific model of the walker, for example, a pendulum model of the legs, to then match the walker to it. Model-free methods, however, do not have an explicit model but rather use the entire capture of the gait to perform the recognition, for example by averaging the silhouette of the walker over time as a gait energy image. Accelerometer-based systems also average the gait into a feature representation either by segmenting the gait into its gait cycles or by using frames with a fixed size.

Utility

Gait recordings are important for medical diagnosis of gait abnormalities [165]. Another more casual example would be the recording of the gait pattern to count the steps a person has performed during a day [336]. Further, gait patterns are often recorded in videos unintentionally, for example when people walk in the background of a recording. Here, the utility of the gait is to appear natural and convincing to the viewer of the video [144].

Threat space

Due to its omnipresence in everyday life, human gait is easy to capture, especially because most capturing methods are unintrusive and do not require the participation of the victim. Additionally, it has been shown that gait recognition is very robust to video quality and obfuscation making it very much suited for surveillance systems [369]. Besides identifying humans it has also been shown that gait can be used to infer private attributes like gender [288]. Considering all this the threat to gait biometrics is already large. What's more, with recent developments in richer capturing methods such as LiDAR [96] or cheap motion capture suits, it is to be expected that the threat space for gait will even increase in the coming years.

Anonymization Techniques

In the following, we present the gait anonymization methods found in the literature, sorted by our taxonomy.

Random Perturbation Hoang et al. [128] propose a fuzzy commitment scheme based on Bose–Chaudhuri–Hocquenghem (BCH) codes for storing accelerometer gait templates. After the feature extraction and binarization of the accelerometer data the reliable bits are extracted. These bits are then XORed with the BCH encoded secret key to gain the secure γ. Additionally to the γ, the hash of the secret key and some helper data are stored. During the authentication phase, the extracted reliable bits are XORed with the secure γ and then decoded with BCH. The result can then be hashed and compared to the hash of the secret key. While the false accept rate is promising the false reject rate of this scheme must be improved to be more user friendly.

Noise Injection The influence of noise injection on the performance of accelerometer/gyroscope authentication systems was studied by Matovu et al. [219]. For their approach, they generate a time series of noise values drawn from a uniform distribution and then merge the original time series with the generated one.

A noise injection approach for gait in videos was developed by Tieu et al. [352]. They use a convolutional neural network (CNN) to mix the gait of a second person (noise gait) into the original gait. In the first step, the silhouette for both the original and noise gait is extracted from a black and white representation of the input videos. The noise gait is selected hereby to have the same size and view angle as the original gait to achieve a more natural result. The silhouettes are then fed into the CNN which uses shared weights networks to abstract them and then merges the abstracted representations via a third network. In a post-processing step, the original gait is replaced with the newly merged gait. Depending on the view angle they achieve identification rates between 20% and 1%. The authors further improve their method in a follow up paper [353]. Here the noise gait is generated via a generative adversarial network (GAN) that takes Gaussian noise as input and outputs noise silhouette. Instead of using a CNN they then use a self-growing and pruning GAN (SP-GAN) to fuse the noise and original gait. Here the identification accuracy was between 30% and 10%. Further, they propose an approach to colorize the resulting black and white silhouette [354]. Hanisch et al. [115] investigated multiple anonymization techniques to protect identity and gender of walkers recorded via motion capture suits. One of their techniques was to add Laplace noise to all body positions of the walker, however their results show that effectively anonymizing was not possible without destroying the utility (measured as naturalness via a user study). Another paper that performs simple noise injection is by Meng et al. [231], they also show that the noise level required for effective anonymization destroys the utility of the data.

Coarsening Nair et al. [252] experiment with coarsening the frame rate, positional accuracy, and dimensionality of VR motion data. They find that while these techniques can reduce identification rates for individual motion sequences, they do not allow effective anonymization on a per-session basis and are therefore not effective for anonymizing motion data.

Feature Removal A feature removal approach for privacy-preserving activity recognition via accelerometers is proposed by Jourdan et al. [154]. They extract various temporal and frequency features from the accelerometer data such as mean, correlation, energy, or entropy. Via experiments, they then determine the influence of each feature for activity and identity recognition. They find that the temporal features contribute more to identity recognition and frequency features more to activity recognition, therefore they remove the temporal features. Their results show a good trade-off between activity recognition (96% reduced to 87%) and identification (90% reduced to 40%). Debs et al. [69] do a similar simple feature removal approach, but they first transform the signal using a short-time Fourier transformation before randomly removing 10% to 90% of the data. Garofalo et al. [99] propose a temporal convolutional network as feature extractor which is trained via adversarial training. After the feature extractor created a feature vector it is evaluated by an identity verifier and an attribute classifier which results are then used as the loss function for the feature extractor training. Rouge et al. [313] developed an anonymization technique for accelerometer motion data. Their technique is to first extract appropriate features from the raw data using a short-time Fourier transform. They then train a random forest classifier to perform action and identity recognition. Using the trained random forest model, they then determine the importance of the features for both classification tasks and remove those that are only important for identification.

Another technique tested by Hanisch et al. [115] was to remove body parts from gait motion capture data to see their impact on the recognition of identity and gender. They found that the gait data is very redundant and even when only the data for the head is kept identification success remains close to 60%.

Continuous Conversion A continuous conversion approach is blurring, in which persons in videos, including their gait, should be de-identified. As a first step, the silhouettes of the persons in the videos are tracked and segmented to then apply the blur. Agrawal et al. [9] proposed two blurring approaches exponential blur and line integral convolution (LIC). Exponential blur regards the video as a 3D space with the time as the z-axis and then calculates a weighted average of the neighbors of each voxel to blur via an exponential function. LIC works with the bounding box of the walker silhouette and maps it onto a vector field which is then used to calculate the output pixels. To counter reversal attacks against the blur randomization of the blurring functions at each pixel is proposed. Another blurring approach is proposed by Ivasic-Kos et al. [144]. They apply a Gaussian filter to blur the silhouettes of walkers. The filter calculates a weighted average of the color of the neighboring pixels, with the weights decreasing monotonically from the central pixel.

Halder et al. [112] work on gait anonymization in videos. They first extract the gait silhouettes from a large number of videos. They then perform a k-means approach to cluster the silhouettes to generate a database of key gait poses. To anonymize a given gait sequence, they also extract the gait silhouette and match it to the closest key pose in the database. The key pose sequence is then used to generate a new video sequence using a GAN. Their evaluation results show that their approach only slightly reduces the identification rate against multiple recognition systems.

Moon et al. [244] investigate the use of adversarial training for anonymizing motion data. They train different machine learning models on 3D pose data to maximize action recognition while minimizing the identification. Their evaluation on the ETRI-activity [147] and NTU60 [188] datasets shows that they can achieve both a high utility for the action recognition and identification rates close to chance. Nair et al. [251] also propose an adversarial approach for the anonymization of VR motion data using a Siamese architecture for the training of the anonymization. Instead of using only the motion sequence as input they also add a random vector. As before they train their model to achieve good action recognition and low identification.

Thapar et al. [350] consider the anonymization of gait in egocentric videos, which are videos that are recorded from a first-person perspective. They first learn the identities of gallery videos via the rotation of the camera which is then transformed into the camera rotation signature via guided backpropagation. This camera signature is then applied to the target video, mixing the gallery identity and the target identity. In their evaluation they test the identification of persons and find that the EER increases from around 20% to around 50% while the activity recognition is reduced by about 10%.

Continuous Conversion + Discrete Conversion An approach that combines both continuous and discrete conversions for walkers in videos is proposed by Hirose et al. [127]. First, they extract the silhouette and the gait cycle of the walker. The silhouette is then transformed via a deconvolutional neural network encoder into a silhouette code. The code is converted by using a k-same approach in which the k-nearest neighbors of the input code are selected and then a weighted average is computed. The gait cycle is transformed via a continuous, differentiable, and monotonically increasing function. In the last step, the new video is generated by feeding the perturbed silhouette code and gait cycle into the convolutional neural network decoder. Their evolution shows that the gait recognition drops from about 100% down to 29%, 21%, and 4% depending on the recognition model.

Evaluation

Gait de-identification is evaluated in the literature via gait recognition systems or human observers with the recognition accuracy as the main metric, but there are also usages of the F1 score, equal error rate (EER), or false acceptance rate (FAR). To access the utility loss there is a larger variety of metrics, usually to either quantify the naturalness of the de-identified gait or to perform another kind of recognition, such as activity. One specific evaluation method we observed was by Matovu et al. [219] in which the authors used the biometric menagerie to observe the de-identification influence on different types of users in biometric authentication systems.

3.4.2. Handmotions

We use the term hand motions as an umbrella for all hand motion related biometric traits, including handwriting, keystrokes, mouse movements, and hand gestures. These traits mostly differ by how they are recorded and what kind of hand motions are performed. Handwriting can be captured offline or online, depending on if only the resulting written text or a real-time capturing of the hand while writing is being used. For this survey, we only consider the uniqueness of one writing style and not the linguistic style (Stylometry) of the written text. In modern life, handwriting has been mostly replaced by typing on keyboards which also is an important biometric factor as individuals can be identified by the timings of their key presses. Besides keyboards also the usage of computer mice creates unique patterns, as their trajectories and clicks are again a biometric factor. Lastly, hand motions can be directly captured using optical or accelerometer tracking techniques.

Hand motion recognition encompasses multiple recognition techniques for different capture modalities, here we give an overview of handwriting, mouse movements, keystrokes, and gestures. For handwriting bases hand motion recognition the input handwriting sequence is often adjusted for its baseline, scaled to a normal writing style, and segmented to meet the demands of the classifier [283]. Handwriting is further dependent if it was captured while the person was writing (online handwriting), for example with a digital pen, or only handwriting itself is captured after the person has finished (offline handwriting). The recognition for mouse movements relies on the trajectory, speed, single, and double clicks performed with a mouse as features. Keystroke-based hand motion recognition is based primarily on the timing differences between key up, down, and hold events. Besides individual events, the differences between two successive events or even three successive events are also used as features [408]. Hand motion recognition via gestures can be split into 2D gestures which are performed on a flat surface (e.g. on a smartphone) and 3D gestures which are performed in mid-air. Sherman et al. [330] use the trajectories of each finger and first resamples them using a cubic spline interpolation to get a lower sampling rate, removing unwanted jitter. To calculate the distance between two gestures dynamic time warping is employed with various distance metrics.

Utility

The utility range for hand motions is large and diverse. For handwriting the resulting text must be readable either by humans or computers, the particular handwriting style is usually not important. This is different for signatures, as their main purpose is to facilitate the identification and verification of the signer's identity, hence their particular style is important, while the readability of the name is less important. Since the other hand motions mostly serve as input modalities for computer systems their utility as input modality [401] must be kept precise and timely to keep their utility. For hand gestures [320], there is additionally its utility for non-verbal communication.

Threat Space

The threat space for hand motion is diverse as the usage of our hands is unavoidable in most everyday tasks and as we often use digital devices the recording of hand motions happens most of the time without us realizing it. As many studies have shown hand motions can be used to identify individuals by their handwriting [283], keystroke dynamics [13], mouse movements [308], and gestures [387]. Besides identification our hand motions also often convey meaning such as when we write a text on a keyboard, the semantics of hand motions can be sensitive too, such as when we enter passwords or write private messages. Specific medical conditions manifest themselves in hand motions, such as hand tremors in Parkinson's patients [148]. Further, hand motions convey information about our emotional state [339].

Anonymization Techniques

In the following, we present the suitable methods for hand motion anonymization, with the exception of mouse movements as we did not find any suitable papers for it.

Random Perturbation Maiorana et al. [209] propose a template protection method for online handwriting which splits a handwriting sequence into segments and then randomly mixes the segments before convoluting them. The same shuffling approach is taken by Maiti et al. [210] to prevent keystroke inference attacks via wrist-worn accelerometers, however, they do not convolute the segments. The approach was only evaluated with 4 participants. Another study investigating the permutation of keystrokes is performed by Vassallo et al. [367], in their evaluation they only investigate the utility reduction. Goubaru et al. [105] propose a template protection scheme for online handwriting templates. They extract the pattern ID for a user by using a common template. The pattern ID is then XORed with a secret that was encoded by an error-correcting code. The result is stored as the template. For the verification, the pattern ID is again extracted and then XORed with the template.

Noise Injection Migdal et al. [237] add delays to keystroke timings. Shahid et al. [325] propose to use the Laplace mechanism on the 2D coordinates of handwritten text to achieve local differential privacy.

Coarsening Vassallo et al. [367] explore suppression of keystrokes to preserve the content of the typed text in a continuous authentication scenario. Maiti et al. [210] also focus on keystrokes privacy and propose two coarsening methods to prevent keystroke inference attacks via wrist-worn accelerometers. In their first approach, they simply detect if a user is typing via several features and then block the access to the accelerometer data to prevent attacks. Their second method reduces the sampling rate of the accelerometer.

Discrete Conversion For discrete conversion we found the following techniques aimed at template protection. An online handwriting template protection scheme is proposed by Sae-Bae et al. [317] which decomposes signatures into histograms on which the authentication is performed. They use one-dimensional histograms to capture the distribution of single features and two-dimensional histograms to capture the dependence between two features. Migdal et al. [238] propose a template protection scheme for multiple modalities, including keystrokes. Their scheme combines multiple pieces of information, such as ip addresses, with the keystroke information and then computes a biohash on it. Leinonen et al. [180] investigate the anonymization of keystroke timing data using two rounding approaches which effectively sort the timings into buckets. Their approach appears to be effective as the identification drops from close to 100% to below 10%. Vassallo et al. [367] explore substitution of keys with a random nearby key to preserve the content of the typed text in a continuous authentication scenario.

Figueiredo et al. [92] have developed a modeling language that can be used to design new gestures for applications. The gestures can then be recognized on the recording hardware, eliminating the need to give the application access to the clear data. No privacy evaluation was performed. For privacy friendly gesture recognition Mukojima et al. [249] designed a system which illuminates the hand with a random pixel pattern and captures the remaining light on the opposite site of the hand with a detector. From this reduce data collection the shape of the hand is reconstructed via machine learning. The authors did not evaluate the privacy protection of their approach.

Continuous Conversion Maiorana et al. [209] propose two continuous conversions for online handwriting templates: A baseline conversion which first splits a handwriting sequence into multiple segments based on a secret key and then convolutes the segments. And a shifting transformation that applies a shift to the initial sequence. The template matching is performed on the protected template. For the anonymization of gestures which have been captured via inertia measurement unit (IMU) sensors Malekzadeh et al. [213] propose two separate auto encoders. The first auto encoder is supposed to replace sequences in the data which have been classified as sensitive with a generated neutral sequence. While the second one should minimize the mutual information between the data and the identity of the user. Their approach reduces the identification from 96% accuracy down to 7%. Fan et al. [87] also propose using two encoders, they use one for task encoding and one for identity encoding and then feed both encodings into the decoder. This system is trained in an adversarial approach to reduce identity recognition and increase action recognition using a small sEMG dataset.

Another auto encoder based approach is proposed by Saunder et al. [320] in which the sign language motions of one person are transferred onto another one. Their technique is two fold they first extract the pose of the source video and encode this to a set of pose features. Secondly they encode the style of the target appearance using an appearance distribution. The encoded pose and style are then combined to generate a new image. It was not evaluated if the persons can be identified by their hand motions only. A second approach to perform sign language anonymization was proposed by Xia et al. [380]. They use an estimation of the motion regions and then use optical flow in combination with a confidence map to encode the motions of the source and driving video. Then the anonymized video is generated via an auto encoder from the source video, optical flow and confidence map. To keep the utility of the sign language high they use a loss function which especially focus on the difference between hand and face motion of the driving and anonymized video. Again no evaluation if the persons can be identified by their hand motions was performed.

Evaluation

Hand motion anonymization is mostly evaluated in the context of authentication and as such the false positive rate (FPR), false negative rate (FNR), and equal error rate (EER) are important metrics for evaluating the performance. But there is also the usage of recognition approaches for the evaluation which uses the accuracy of identity, age, gender, and handedness inference. A unique evaluation approach we found was used by Goubaru et al. [105] who used the randomness of the template bits via occurrences and autocorrelation to evaluate their approach. Again we find that more critical evaluation approaches are possible, as the EER will most likely overestimate the anonymization performance as it tries to achieve a low false positive rate.

3.5. Discussion

All reviewed behavioral biometric traits have in common that they are captured as a time-series tracking the change of the trait over time. Most traits, such as gait, hand motions, voice, and eye gaze are overt traits that can be observed from a distance and do not require the participation of the subject. These traits are often captured as a byproduct for other recordings, for example, video recordings. EEG and ECG on the other hand are secret traits that can mostly only be recorded by directly attaching sensors to the subject to measure them. We found the most anonymization methods for voice and the least for EEG. For the traits touch, thermal, and lip-facial we could not find any mechanisms.

The **utility** of these traits is very diverse and is mostly unique to each trait and the application using it. It ranges from utilities such as the naturalness of a motion to the intelligibility of utterances.

Regarding their **threat space**, the traits are similar to each other, as due to the pervasiveness of digital capturing devices more instances of them are captured. Wearables and mobile devices are of especial interest as they are attached to the subject and can therefore allow continuous capture of behavioral data. As our literature review has shown all traits can be used for both identity and attribute inference, which then can be abused for a wide variety of privacy threats such as surveillance, identity theft, or private attribute inference. The privacy goals, identity protection, and attribute protection are also the same for all the traits. However, voice has an additional privacy goal in which the content of the speech should be made unintelligible.

For the **techniques** (see Table A.2 and Table A.3) that we reviewed, we found that most of them fall into the category of continuous conversion, followed by feature removal and noise injection. Next are random perturbation and discrete conversion, with most discrete conversion methods aiming at template protection. Coarsening is the category with the least amount of methods. We observe several differences for the categories of our taxonomy, for the removal methods we find that the removal is not directly reversible, however, due to the high redundancy in behavioral biometric data it still might be possible to reconstruct the removed data. For the conversion methods, we often observe that the parameter space for the anonymizations is often rather small, making it possible that an attacker can link clear and anonymized data by brute forcing the parameters when the anonymization technique is known. In general, we find that the reversibility of conversion techniques still has to be evaluated better. For noise injection techniques we find that the strong dependency both temporal and physiological is a problem since they can be used to filter out the noise.

With regard to the techniques providing **differential privacy**, we have observed that none of them can be used continuously over time without completely compromising user privacy. The reason lies in that the privacy budget is necessarily finite, which means, by the sequential composition property of differential privacy [226], that it will be consumed completely at some time instant. Surprisingly, this appears to be in contradiction to the intended use of most of the applications where differential privacy is guaranteed, namely, continuous monitoring in healthcare scenarios, and identification and authentication services (which clearly are not single-use services). In that respect, the use of related privacy notions intended for continuous observations (e.g., w-event differential privacy [159]) may come in handy. In general, more research is needed on how to effectively apply differential privacy to behavioral data.

We made the observation that most methods do not **manipulate the temporal aspect** of their data. Notable exceptions are Hirose et al. [127] and Maiti et al. [210]. Since all traits result in time series data manipulating the temporal order or time differences between events could lead to some general anonymization techniques which work for multiple traits. For attribute protection we find anonymizing intrinsic attributes (e.g., age, sex) to be difficult as it is not clear which part of the behavioral data is relevant for these attributes. We therefore find generative machine learning approaches a promising approach to address this problem as the machine learning models can learn the intrinsic dependencies between data and attributes. Further, we noticed a lack of even basic understanding of **users' privacy awareness** and concerns about behavioral privacy. These are necessary to design protection techniques that consider user needs and requirements.

We found that the **evaluation methodology** between the traits and methods is rather similar. In general, an inference/recognition system is being used on the clear and on the anonymized data and then the difference in accuracy is reported, often without retraining the inference system on the anonymized data. We find this methodology too simple as the underlying assumption is that the attacker is not aware of the anonymization. A notable exception are more recent voice anonymization techniques which now mostly rely on the benchmarking framework of the VoicePrivacy Challenge to evaluate the privacy and utility of their techniques. This shows that community initiatives can provide a common basis for comparison and improve the overall evaluation methodology of a field.

Only a small number of articles compare their own methods to that of others, and due to the differences in attacker models and data sources, they are difficult to compare for the readers. We also found that there are not many approaches [402, 299] to formalize the privacy of behavioral biometric anonymization methods and most of the evaluations rely on empirical privacy estimations. Another problem is that the evaluation methodology is too close to the recognition system evaluation methodology which seeks to infer persons in a large dataset with poor data quality, while an anonymization method should also work on a small group size with high data quality. To reliably estimate the level of protection that anonymization provides, we believe it is important to assume the worst-case scenario, in which identifying individuals is easy and anonymization is most difficult. We believe that the lack of available datasets is one of the main problems which keeps the less researched behavioral biometric traits back. For possible future work, we see the anonymization of eye-gaze and motion data as promising areas of research, as many challenges remain, like achieving good utility and real-time applicability. Similar to the VoicePrivacy Challenge, most behavioral biometrics would benefit from community-driven evaluation frameworks to increase the comparability and rigor of privacy and utility evaluations. One area where many behavioral biometric traits are combined is the creation of digital twins, where it is an open question whether anonymizing the behavioral traits independently of each other is sufficient to create privacy-friendly digital twins, e.g. for mixed reality.

3.6. Problem Statement

Our survey on anonymizing behavioral biometric data revealed that, although there are many different behavioral traits, there are also many similarities in the problems encountered and their solutions. In this thesis, we address the most pressing problems we identified in the anonymization of motion data. Below, we provide an overview of the main issues we will address.

- The privacy issues associated with motion data have received some initial attention, as evidenced by the many gait recognition approaches that have been developed. However, these approaches only require identification to work; a deeper understanding of why it works is unnecessary. Anonymization is different in that we only want to remove personally identifiable information from the data while leaving the rest intact.

Therefore, we will study which features allow individuals to be identified from gait motion data.

- Advances in mixed reality technology have made more types of motion tracking possible. One new type of motion tracking that has become available in recent years is tracking facial expressions. Mixed reality headsets capture their users' facial expressions and transform them into abstract representations, which are then used to animate digital avatars' faces. While it is already established that gait can be used to identify individuals, this remains an open question for facial motion data. Therefore, we study the privacy issues posed by facial motion data.
- Our comparison of evaluation methodologies for behavioral biometric anonymizations showed us that the same methodology can be used to evaluate the privacy of anonymizations across traits. However, we found that this methodology is not up to par as most evaluations assume weak attackers who are unaware of the anonymization process. This can result in unreliable reports on the effectiveness of the evaluated anonymizations. Therefore, we work on improving the methodology for evaluating biometric data anonymization.
- An important precondition for conducting meaningful research on anonymizing human motion data is having access to the data. As our survey showed, most current human motion data is video data, which is not suitable for anonymization research because it requires larger sample sizes per person. Furthermore, with the rise of mixed reality, motion tracking is becoming more detailed. This results in people being captured as 3D motion captures, rather than from one perspective as in 2D video. Since datasets can stimulate new research in their respective fields, we aim to collect new and novel motion dataset to develop anonymization techniques.
- Although there are many anonymization methods for behavioral biometric traits, most of these solutions focus on voice data, and few focus on motion data. Anonymizing motion data is a complex task because identification from motion data is very robust, and identifiable information is contained in almost all aspects of the data. Therefore, we will pursue an anonymization approach that considers all dependencies included in motion data.

4. Investigating the Privacy Issues of Motion Data

Before we start working towards developing an anonymization to protect motion data from privacy inferences we must first understand what makes the privacy inferences possible. In this chapter, we seek to investigate the problem of person identification through motion data. For this we specifically look at human gait (i.e. human walking motion), as it is one of the most identifiable human motions and Examples from China[1] show that gait is very much suited to be used for surveillance purposes alongside face recognition. Further, there is a wide variety of techniques to capture human gait via sensors.

When compared to person identification via faces, gait has advantages as it can be done from distances at which the face is not yet recognizable or occluded by objects such as face masks. It is believed that distinguishing individuals from afar was an important human survival mechanism in the past, as it allowed to recognize if an individual was a friend or foe before the person was close enough to be a potential threat [395].

As a starting point for finding categories of features for our computational experiments, we look at the literature on human gait perception. This line of research has long been a topic of cognitive science and investigates how humans identify other humans by their gait (e.g., [151, 361, 59]). We would like to emphasize that we are not interested in designing a novel attack, and we are less interested in investigating the robustness of specific anonymization schemes. Instead, we are interested in the question of which features in precise gait data yield identity or attribute disclosure of the individuals, if they coincide with those known from psychological research on gait and person perception—and to which extent they are inter-dependent and hence cannot independently be suppressed/perturbed for anonymization.

For a systematic analysis we use machine learning (ML) to get an estimate of how much identification potential gait data has. We then design perturbations for each of the feature categories that remove specific features in the gait data, to then measure how much the recognition performance drops. Where possible, we try to manipulate a feature alone, so that we can estimate how much identifying information the feature contains, as well as how much it shares with other features. To establish how much utility is retained after anonymization, we perform a user study in which the participants rate the naturalness of the resulting gaits. Key contributions of this chapter are as follows:

[1] `https://apnews.com/article/bf75dd1c26c947b7826d270a16e2658a` [apnews], accessed: 17.08.2022

This chapter is based on the contribution:

- **Simon Hanisch**, Evelyn Muschter, Admantini Hatzipanayioti, Shu-Chen Li, and Thorsten Strufe. “Understanding Person Identification Through Gait”. In: Proceedings on Privacy Enhancing Technologies. 2023. DOI: 10.56553/popets-2023-0011.

- We categorize human gait features using categories extracted from the literature of cognitive science.
- We systematic study feature contribution for gait recognition for both identification and sex classification.
- We propose simple gait perturbation techniques.
- We presented a utility evaluation via perceived naturalness of the anonymized gait.

4.1. Background and Related Work

In the following section we provide background on current motion capturing including the human and automatic recognition tasks based on gait, and the state of the art with regards to anonymization and explainability-based analyses of identifying features in gait.

4.1.1. Motion Analysis

Humans can recognize and identify biological traits visually through the use of static information such as shape or other cues. Biological motion is one additional important factor. Human newborns, infant monkeys, and even freshly hatched, visually naïve chicks show a preference for motion cues that move in motion patterns suggesting animacy [198].

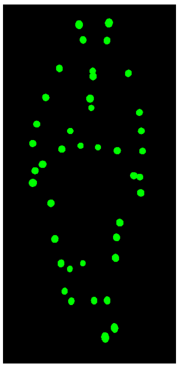


Figure 4.1.: Example of a human walker represented as a point-light display (PLD).

An established and reliable method of investigating biological motion processing without other potential sensory information or cues (e.g., color, texture, or form-based features such as facial configurations, hairstyle, or clothing) is the use of *point-light displays* (PLDs) [151] (see Fig. 4.1). This method of using impoverished moving dot displays helps to isolate motion information from other cues. Thereby, a small number of dots represent the head and major joints of a human body in various scenarios such as social interactions [28, 214] or—as the focus of this work—during gait [361]. Indeed, many identifying features of a human walker can be recognized from such PLDs.

Gait analysis is the study of human locomotion (walking and running) and defines walking as a series of gait cycles. A **gait cycle (stride)** is the period when one foot contacts the ground to when that same foot contacts the ground again (see Fig. 4.2). Each gait cycle has two phases: the stance phase, when the foot is in contact with the ground; and

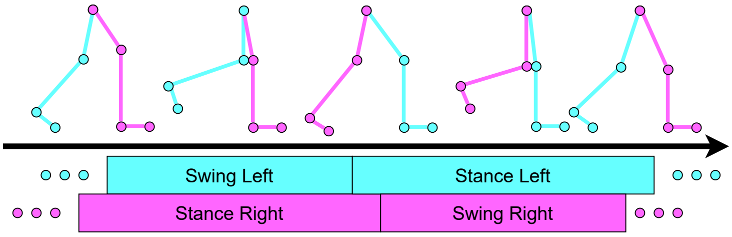


Figure 4.2.: A schematic representation of the gait cycle.

the swing phase when the foot is not in contact with the ground. In vision-based gait analysis such as PLDs, kinematic data such as position and velocity are captured. These are used to relate motion parameters, such as joint angles and joint velocity, with qualitative gait parameters, such as step length, walking speed, the pace and rhythm of steps, stance and swing times, as well as arm swing, vertical head movement, pelvic rotation and the extension and flexion of the limbs and shoulders. As walking consists of a series of multiple gait cycles, gait data typically also contains fluctuations. These are small variations such as asymmetry and variability in step and stance time, step velocity, or step length [71]. These gait parameters can subsequently be used to extract statistical features (e.g., mean, standard deviation, skewness) for gait analysis.

4.1.2. Human Gait Perception

Human observers have no trouble making sense of the very limited information presented through PLDs disconnected dots, representing actions such as the specific categorical biological motion content [395, 268] of human gait (walking or running). Research has suggested that humans are especially tuned to recognizing conspecifics and this preference is likely to already emerge in the visual system. Psychophysical and neuroscientific studies have shown that at least two processes play a role in person perception (here defined as the recognition of human bodies and their biometric features based on vision).

On the one hand, form-from-motion cues [332, 395] are cues that are rooted in basic perceptual abilities to see structure from motion. That is, the shape and form of an object or person are revealed more clearly through motion. The human visual system benefits from the motion direction information in order to extrapolate the overall shape of an object or person. These cues provide time-invariant information about body form by enhancing the shape presentation of a person [395] and are susceptible to violations of the hierarchical body form structure [41, 279] as well as to inversion effects [97]. That is, inverting a body in the image plane (i.e., placing it upside-down) results in perceptual impairments. Previous research has proposed that inversion is deleterious to normal human whole-body perception, causing observers instead to rely on local part-based visual features [97]. Evidence for first order configural processing has shown that visual perception of bodies is mediated by spatial configurations of body parts, such as the general body layout (e.g., legs attached to the hip, arms attached to the shoulders), and thus providing intact spatial configurations of bodies [41].

On the other hand, dynamic identity signature [332, 395], describes the idiosyncratic motion pattern of an individual. These features describe the change over time during a walking cycle and rely on nuanced, person-specific motion variations (e.g., the way Charlie

Chaplin walks). Furthermore, research has provided evidence for a two-stream processing of biological motion perception in the brain. That is, biological motion perception relies on both dynamic and static features through motion processing in the dorsal pathway (i.e., area V5 of visual cortex in the brain) in combination with bodily form and appearance information in the ventral pathway in the brain (see Peng et al. [279] for further details on visual recognition of biological motion). In addition to action recognition, human observers are able to identify soft biometric features of actors in PLDs, including sex [242, 97, 170], age [409, 242], weight [247], height [247], handedness [191], in addition to attractiveness [242, 247], identity [62], emotions [242, 153] and causal intentions [214]. Specifically, Kozlowski and Cutting [170] showed that the biomechanical factor center of moment, which is derived from the relative movement of both – the shoulders and hips, plays a crucial role in sex perception in PLDs.

Finally, person recognition depends on familiarity and might take time for the human observer to learn, but is useful for recognizing a familiar person from a distance [62, 361]. Studies have shown that human observers are more sensitive to PLDs of themselves and friends [62, 200], or could learn to identify a small number of individuals based on their motion [361]. Guided by these insights, we aim to investigate if the removal of the certain features will reduce recognition rates, and to which extent.

Specifically, we focus on the features that are easy to extract from existing data sets. Namely, macro and micro features (i.e., statistical features, see Section 4.1.1), perturbations of intact bodies in natural spatial configurations, as well as dynamic (i.e., temporal information) and static (i.e., structural) features.

4.1.3. Automatic Gait Analysis

Current human movement analyses are based on biometric measurements and motions. They are captured vision-based, via pressure plates, or using wearables with integrated inertial measurement units (IMUs). A gait cycle is thereby composed of a chain of individual 2D (video) or 3D (optical marker/IMU tracking) samples at each given time point (pose).

Gait recognition using machine learning models is most commonly based on video data [369]. Video, providing rich information about subjects, facilitates high recognition rates and hence is frequently used for surveillance purposes [26].

Also explicit motion capturing frequently uses video: High-quality vision-based motion capturing uses specialized cameras to track reflective markers on the subject's body. The position of these markers is later reconstructed into 3D position time series, converted into joint angles as a function of time, and subsequently analyzed according to specific research or clinical needs. However, recently, approaches of using a single commodity camera in combination with keypoint detection algorithms and neural networks (e.g., Open Pose or DeepLabCut) have generated convincing results [163, 228, 217].

Gait recognition is also possible based on motion capturing data [25]. Indeed, even simple kinematic features obtained from IMU systems (e.g., position, velocity, and acceleration-based features) or kinetic data from force plates and electromyography (e.g., ground or muscle force parameter) have been shown to yield high recognition rates (see [59] for an overview).

Anonymizing individuals in video surveillance footage for multiple moving object detection and tracking algorithm (e.g., human action tracking) by representing their bodies as simplified objects such as PLDs thus cannot protect their identities. Further, gait can also be used to infer personal attributes like sex [396] and age [409, 85]. Being interested in those gait features that carry information for identification and attribute disclosure of individuals, in the present work we rely on marker-based motion capture data as it is considered the gold standard in the field.

4.2. Methods

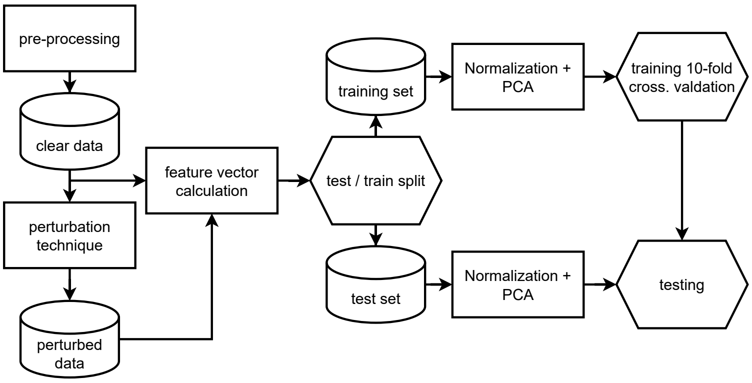


Figure 4.3.: The full data processing pipeline.

The question we sought to answer is how much specific features in the data contribute to the overall gait recognition performance of identity and sex using machine learning. Our overall approach was to first train & test a gait recognition system for each of the recognition goals on clear data to obtain baseline accuracy. Next we obfuscated a feature at a time in the data by either perturbing or removing it to investigate its impact on anonymization. We then repeated the training & testing process and report the resulting recognition performance. The difference in recognition between baseline and perturbed data gives us an approximation of the unique amount a feature contributes to the overall recognition performance. Further, we also measured each feature independently from the other features. However, this is only possible for features we can remove from the data and not for features we only perturb. The full process is shown in Fig. 4.3.

In the following, we provide details about the data set, applied feature perturbations, the implementation of the recognition system, and utility evaluation.

4.2.1. Data Set

As our main goal was to understand the important features of human gait, we chose the highest quality of gait data and used optical 3D marker-based motion capture data for our experiments. This data is considered the gold standard for motion capturing and is recorded using multiple infrared cameras which capture markers on the anatomic landmarks of participants. The benefit of the 3D representation is that there is no dependency on the recording angle like in video recordings. The data consists of multiple samples per participant which are a time series of poses. Each pose contains the 3-dimensional coordinates of each marker (placed on the participant) at a given point in time (i.e., PLDs). The data is also more appropriate for our purpose, as we focus on gait features in the absence of potential additional information (e.g. video recordings).

We used the open-source data set by Horst et al. [133, 134] which consists of full-body kinematic and kinetic data of 57 individuals (29 female, 28 male; 23.1 ± 2.7 years; 1.74 ± 0.10m; 67.9 ± 11.3kg). An optical motion capture system and a full body marker set (62 markers corresponding to anatomical landmarks), as well as two force plates, recorded self-paced walking trials at 250Hz (motion capture) and 1000Hz (force plates). For each participant 20 samples containing a full gait cycle have been recorded (for further details on the data acquisition protocol see Horst et al. [134]).

4.2.2. Data Pre-processing

Our biometric recognition system requires gait sequences to contain exactly one gait cycle. However, the raw sequences in the dataset start at different stages of the gait cycle. Therefore, we pre-process the sequences.

Following the methodology of Horst et al. [134], we trimmed the gait samples to contain only a single stride by using the kinetic force signals of the force plates, using a ground force threshold of 20N. This way all samples are aligned and start at the same point in the gait cycle. The data was then normalized, in order to obtain an equal number of poses for each individual, by resampling each sample to 100 frames. Each frame represents one discrete pose of the individual while walking, the 100 poses then constitute one stride.

4.2.3. Retained and Masked Features

We will now explain the retained and masked features for our experiments, as well as their respective categories. We base our feature categories on previous work in gait analysis and human perception as described in Sections 4.1.1 and 4.1.2. The category name always gives the kind of feature we sought to retain, while the perturbation techniques employed are aimed at removing the other features from the data. For each technique, we strove to design an inverse perturbation technique that only removes the specific feature, while keeping all the others (micro vs. macro, dynamic vs. static). This way we sought to understand how much each feature contributes to the overall recognition rate and if it contains information that is unique to this feature. Since there is interdependence between features, some of the features are partially overlapping for example the walking frequency is dependent on the walking speed and the length of the legs. Table 4.1 gives a brief overview, while the used and obfuscated features are described in detail in the following.
Our **macro** features describe the general characteristics of the walker, such as walking speed, general movement trajectories, walking amplitude, the most significant parts of the walker positions, and overall body parts. Its counterparts are the **micro** features which contain the small variations of the trajectories that remain when the overall trajectories are removed, the walker without its walking speed and step length equalized over all walkers, the least significant parts of the walker positions, and individual body parts. Besides macro and micro, we also investigated the dynamic parts of the gait motion. For this we have two contrary feature categories static and dynamic. The **static** features contain the time-invariant features, such as the average pose of the walker, or the first pose of the walker. The **dynamic** features contain the features describing the motion of the walker, including the differences between the recorded poses, and walker where the static frame (body proportions) has been removed. The following section describes the used perturbation techniques for each feature category. The parameter values have been chosen to

match the used data set. In the end, we briefly detail how we combined the perturbation techniques.

Table 4.1.: Used and obfuscated features

	Macro	**Micro**	**Static**	**Dynamic**
definition	*Step length, walking speed, cadence*	*asymmetry and variability in the macro features*	*Shape and general body layout*	*Time course of changes*
Perturbation 1	Remove variations	Remove trajectories	Static pose	Motion extraction
Perturbation 2		Amplitude/ frequency equalization	Resampling	Normalization
Perturbation 3	Coarsening macro	Coarsening micro		
Perturbation 4	Remove body parts	Keep body parts		

We provide a sample video rendering[2] of all perturbation methods.

Macro Features

The macro features keep the overall characteristics of the walker and remove its smaller variations from the data. We used three perturbation techniques for this: remove variations, coarsening macro, and remove body parts.
Remove variations: In order to extract the ideal trajectory from the gait data we removed the small variations that deviate from the ideal trajectory. The ideal trajectory is here calculated by two different methods: either using a moving window on the marker poses and then calculating a rolling average, or an interpolation. The difference between the two is that the rolling average takes all poses in the window to calculate an average, while the interpolation only uses the poses at the edge of the moving window. The moving window size is given as the distance to the pose which is calculated and is either one or three additional pose(s) before and after e.g., spanning three poses in total or spanning seven poses in total, respectively. This strategy follows a similar idea to low-pass filtering, as it retains the main movement but removes detailed deviations.
Coarsening macro: As we were interested in the most significant information of the walker position, we removed the least significant part of each marker position in a pose for all poses. The effect is that the grid on which the walker moves is becoming more coarse. We removed all digits either below the thousandth (1000) or the hundredth digit (100).
Remove body parts: We measured how much´the motion of an individual body part (head, torso, hip, arms, legs) contributes to the overall recognition performance. This was done by removing the body part from the data by setting its marker positions to zero.

Micro Features

The micro features are the counterparts to the macro features. Here we kept the small variations of the gait cycle and the least significant parts of the marker positions.
Remove trajectories: Contrasting remove variations, we removed the ideal marker trajectories from the data by calculating the ideal trajectory as described in remove variation via either rolling average or interpolation with a window size of 1 or 3. The ideal trajectory was then subtracted from the real trajectory, which leaves us with the distances of the

[2] https://github.com/kit-ps/understanding-person-identification-through-gait-popets-2023

ideal marker positions to the real ones. This strategy resembles high-pass filtering, as it removes the main movement and only retains the minor specifics of the current sample.
Coarsening micro: We eliminated the most significant part of the walker positions by removing the most significant parts of each marker position value. We removed all digits above the hundredth (100), tenth (10), or first digit (1) position
Keep body part: We measured how much recognition performance the individual body parts have alone without the rest of the body. All remaining other body parts are set to zero.
Amplitude/Frequency equalization: The walking amplitude and frequency were equalized between all individuals to perturb their influence on the recognition. Informed by previous studies [361], we calculated a gait representation of each individual by using the average pose, the first four components of a principal component analysis (PCA), and a sinus function fit on these components to represent the gait cycle of a person. We then equalized the frequency or amplitude of the fitted sinus function by means of the group-level average.

Static Features

The static features capture the time-invariant features of the walker by removing the dynamic part of the gait motion. We therefore kept the proportions of the walker.
Static pose: We used only an average pose or the first pose of each sample, thus removing the dynamic component of the gait data.
Resampling: We downsampled the data to 10 frames, and therefore removed most of the dynamic content from the data.

Dynamic Features

The dynamic features are the counterpart to the static features and aim to only retain the dynamic part of the motion.
Motion extraction: Instead of using the individual poses, we used their difference (i.e., keeping only the variations between poses) and hence removed the static features.
Normalization: We normalized the static features in a sequence by either normalizing the height axis (y-axis), all axes or normalizing each dimension over the entire sequence of poses.

Combinations of Perturbations

Besides evaluating each of the features alone, we also investigated their combinations. Two perturbation techniques were combined by applying them sequentially to the data. Due to some techniques (first pose, average pose) not returning a time series, not all combinations of methods are possible. As the overall number of combinations is quite high, we focused on representatives of each class of features. We picked those representatives by their anonymization impact on the data.

4.2.4. Recognition System

To test the impact of omitting features from human gait, and hence their contribution to inference, we implemented a gait recognition system to perform closed-set identification.

We opted for a simple gait recognition system that can be quickly trained, since we train the system using perturbed data for each perturbation. We adapt the gait recognition system by Horst et al. [134] using Python 3.8.3 [296], Scikit-learn 0.23.1 [278], and NumPy 1.18.5 [119]. We used two feature vectors to represent a data sample: **flatten** which concatenates all poses of a sample into a single vector, and **reduced angles** which first calculates a reduced representation of 17 markers representing the main body parts and then calculated 10 joint angles from this representation.

Next, the data was split into train (75%) and test (25%) data. Here we differentiated between the identity and sex recognition. For identity recognition, we split the samples for each identity so that we have every identity in both sets. While for sex recognition we split the samples identity-wise, making sure that every identity is only in one of the sets. We did so to make sure that the classifier cannot learn the identity to perform sex recognition. Following the split, we then scaled the data in each set by subtracting the mean and then scaling with the standard deviation before we performed a principal component analysis (PCA) to reduce the dimensions of the samples. As a classifier, we used a support vector machine (SVM) using a radial basis function (RBF) kernel. For the training of the SVM we used 10-fold cross-validation with the train set before we tested the best performing model on the test set. In order to account for the random splitting of the data, we ran the entire process 10 times.

4.2.5. Utility

Besides investigating the identity and sex recognition performance of our features we also sought to understand how much the features contribute to the utility of envisioned applications. As our use case is to transfer the gait motion onto a digital avatar, the goal is to retain as much naturalness in the motion data as possible. In order to measure the corresponding effect, we performed an online survey with 22 human participants (13 male, age: 18–60 years) which we asked to rate the naturalness of the perturbed gait sequences. Participants were shown renderings of two gait sequences for each perturbation in which the walkers (one male and one female walker, individually) were shown from the side 45 degrees rotated around the z-axis towards the camera. The renderings are identical to the example videos we provided in Section 4.2.3. All sequences were shown in random order. The participants then rated on a scale from 1 (worst) to 5 (best) how natural looking the gait sequence appeared to them. The survey data collection is under the umbrella of the project ("Privatsphäre von Körperbewegungen") approved on 30.09.2021 by the ethical committee of KIT and was conducted in accordance with the Declaration of Helsinki. The survey data was collected in anonymized form.

4.3. Results

In this section we present the results of our obfuscation experiments, by reporting the recognition performance of the chosen feature categories. The results for identity and sex recognition in two contrasting feature categories (macro vs. micro, dynamic vs. static) are reported each. Note, that we report the body part removals (body parts vs. rest body) separate from the macro and micro features for easier comparability.

As we conducted recognition experiments and the classified classes (for both identity and sex recognition) have nearly the same number of samples per class, we selected accuracy as our metric. Accuracy is defined as the number of correctly classified samples divided by the number of all classified samples.

In Section 4.2 we described the two feature vectors we used in our recognition system. Since we were interested in how much identifiable information remains in the data after the perturbation has been applied, we always report the values of the best performing feature vector.

4.3.1. Macro vs. Micro Features

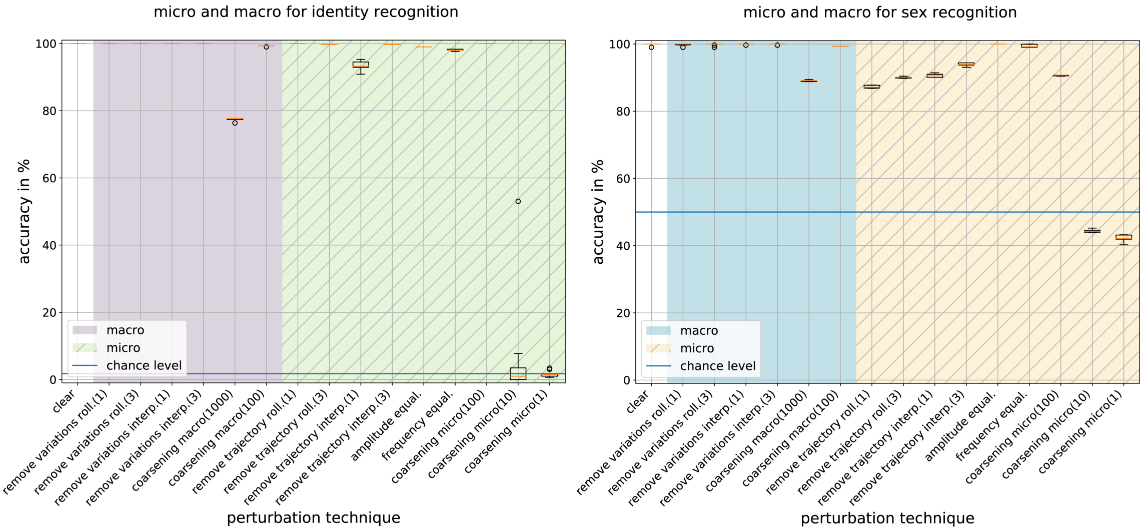


Figure 4.4.: Boxplots of accuracy results for micro and macro features for identity (left) and sex (right) recognition given in percent.

We start by comparing macro to micro features. For both, identity and sex recognition (Fig. 4.4), we can see similar effects for the macro features: The variation removal via rolling average and interpolation shows no effect on the accuracy. The coarsening of all digits below the 100th digit has no effect, while coarsening from the 1000th digit position leads to a drop in accuracy for sex recognition to 91% and identity recognition to 77%. For the micro features, we see a difference between identity and sex recognition. Only trajectory removal using an interpolation window of 1 drops the accuracy of the identity recognition, while all of the others lead to a drop in sex recognition to about 90%. For the micro coarsening methods we again see that identity recognition is not affected by coarsening everything higher than the 100th digit, while for sex recognition we see a drop of accuracy to 90%. Then coarsening the digits above the 10th digit leads for both, identity and sex recognition, to chance level accuracy. The results show that sex recognition is more dependent on the macro feature than on the micro features, while the identity can be perfectly inferred from both of them.

4.3.2. Individual Body Parts in Isolation vs. Reduced Whole Bodies

Next, we evaluate perturbations of individual isolated body parts in contrast to reduced whole body configurations (i.e., certain body parts were removed) (see Fig. 4.5). On the

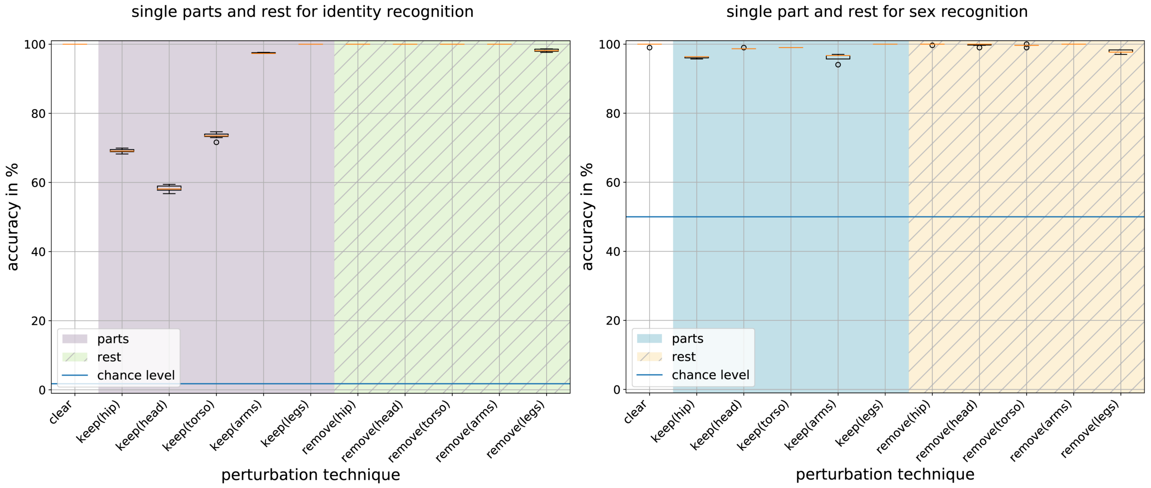


Figure 4.5.: Accuracy boxplots of results for individual body parts and all of the remaining body parts for identity (left) and sex (right) recognition given in percent.

one hand, only the specified body part is used for the recognition ("keep"), while on the other hand, the whole body minus the specific body part ("remove") is employed. Fig. 4.5 shows that the removal of the legs slightly reduces the identity recognition accuracy to 97%. At the same time, it is the only body part that achieves 100% recognition accuracy alone. In contrast, keeping only the head as the standalone body part achieves the strongest prevention from identity recognition, reducing the accuracy to less than 60%. Only slightly improved performance is achieved by the standalone body parts torso or hip.

For the sex recognition, we find the same small reduction in accuracy for the removal of the legs as we saw for identity recognition, while it is again the only body part to achieve the full recognition accuracy as a standalone body part. However, for the other body parts, we find that their removal does not impact the sex recognition score. Additionally, our data shows only small effects on using only individual body parts in isolation. Comparing identity to sex recognition, head, hip, and torso alone fare much better for sex than for identity recognition. These results suggest that even the limited form information which is integrated over time into dynamic form information is sufficient to identify biological traits such as sex or even identity.

This finding is in line with human perception research. For example, Kozlowski et al. [170] found that longer strides are perceived as more masculine. Center of moment contains sex information (see also Section 4.1.2; [170, 242]). That is, as long as the stimuli contains information about certain body parts, sex and identity recognition is possible.

4.3.3. Dynamic vs. Static Features

Thirdly, we investigate the effects of dynamic and static feature perturbation on recognition performance. In the case of identity recognition, depicted in Fig. 4.6, we observe that only using the average pose or the first pose reduces the recognition accuracy slightly to 91% and 94% respectively, while other feature manipulations show no effect on identity recognition. For sex recognition, our results show that while static features have close to no effect on accuracy, all dynamic features appear to do so. So we can conclude that the

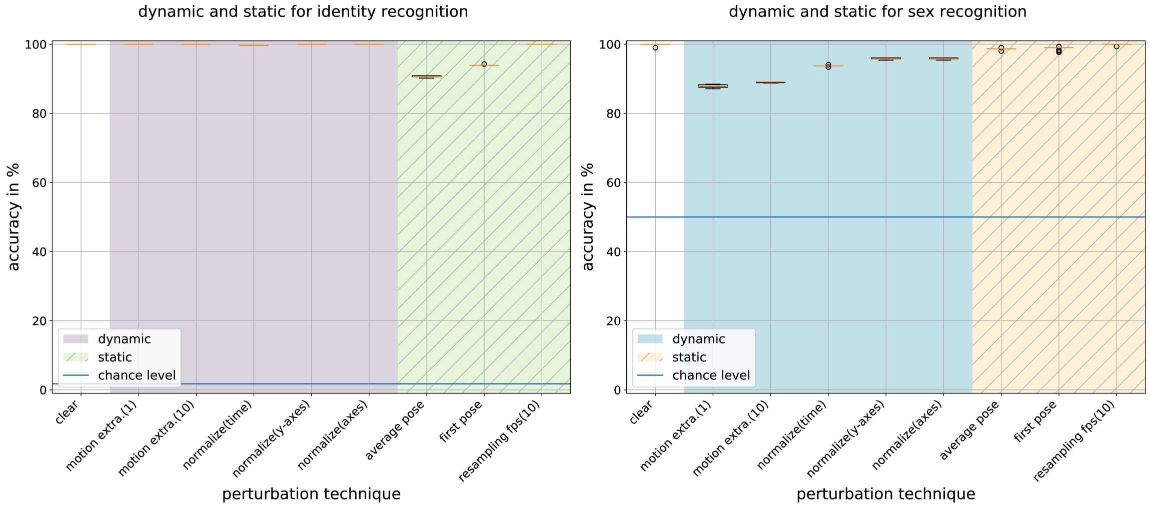


Figure 4.6.: Boxplots of accuracy results for dynamic and static features for identity (left) and sex (right) recognition given in percent.

static features are more important for the sex recognition than the dynamic ones.

4.3.4. Combination of Features

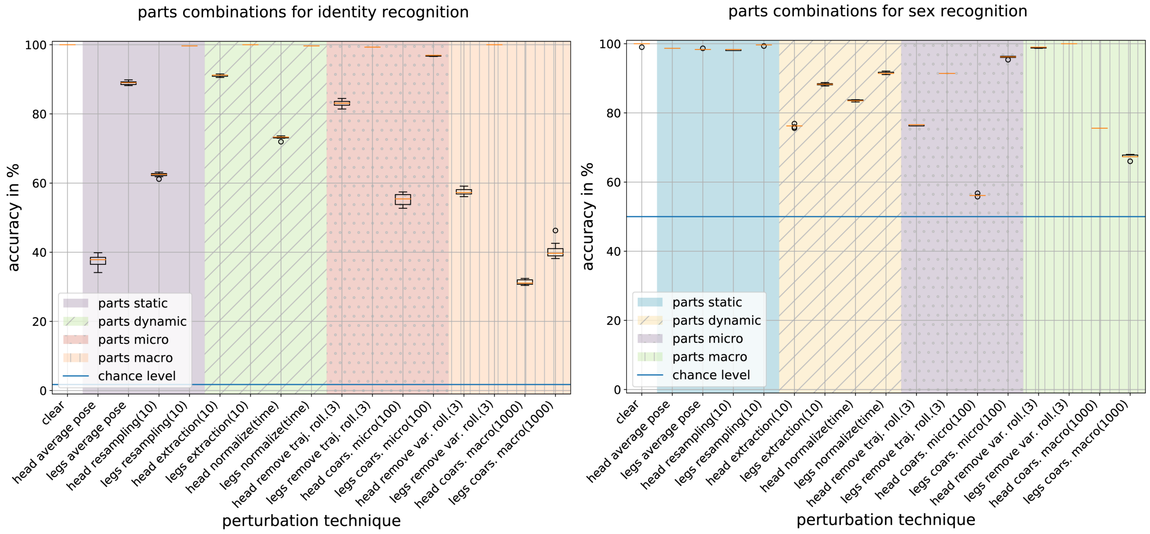


Figure 4.7.: Accuracy boxplots of results for legs and head in combination with the other categories for identity (left) and sex (right) recognition given in percent.

Here, we evaluate the combination of selected perturbation techniques from each category. Due to the further removal of data, we expected to see larger reductions in the classification accuracy for both identity and sex recognition. We also expected that with fewer data available the classification process becomes more unsteady and therefore the variance between the results will be larger. Further, the reduction of data can lead to a simplification of the data, which then is easier to classify. First, we describe the results for body parts alongside other features, before moving on to micro and macro combinations.

The combination of body parts head and legs with the static, dynamic, micro, and macro categories for identity recognition are shown in Fig. 4.7. Most of the legs combinations remain at 100% accuracy. Only in combination with average pose and coarsening micro (100), a slight decrease in accuracy can be observed. When the legs are combined with coarsening macro (1000) we observe a large decrease in accuracy to close to 40%, while both of these perturbations alone do not have an effect on the accuracy. The head (head alone achieves 60% identity recognition) combinations are more of a mixed bag. While average pose and coarsening macro further reduce the accuracy; resampling, coarsening micro, and remove variations do not have an additional effect on the accuracy. However, motion extraction, time normalization, and remove trajectory lead to an increase in the recognition accuracy. All three methods focus more on the smaller variations in the data, providing an indication that the identification of individuals via their head motion is more dependent on the dynamic parts than the general movement.

Focusing on the combinations with head and legs for sex recognition, we find that while there is no effect on accuracy in combination with static features, the combination with dynamic features has deleterious effects on accuracy. Specifically, the combination of head and motion extraction results in a drop of accuracy to 75%. For the micro combinations, we again find that the combinations with the head suffer the largest accuracy reduction. Here the combination with micro coarsening nearly reaches chance level, while the same combination with the legs stays above 90%. When we compare this with the macro coarsening of the macro features, we find that the legs drop to a lower accuracy than the head. This leads us to conclude that the sex recognition via the head data is much more dependent on the macro part of the head position, while the sex recognition via the legs depends more on the micro part of the positions.

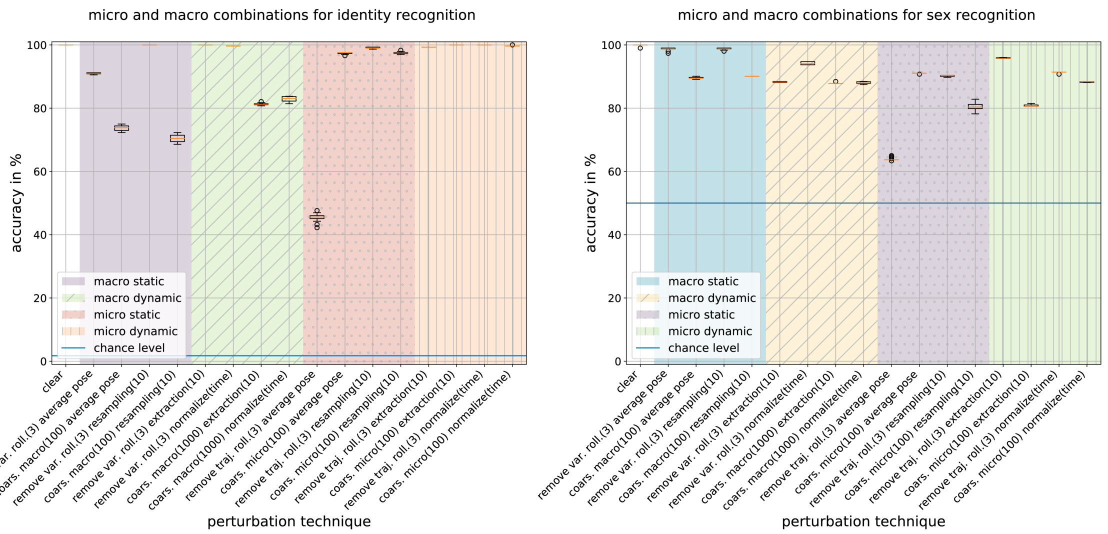


Figure 4.8.: Accuracy boxplots of results for macro, micro, dynamic, and static combinations for identity (left) and sex (right) recognition given in percent.

The results of macro, micro, dynamic, and static feature combinations for identity recognition are shown in Fig. 4.8. The macro-static combinations show accuracy decreases for the combinations that contain macro coarsening. We also see these decreases when we look at the combination of macro dynamic features in which the macro coarsening leads to a decrease in performance. The last combinations show an accuracy decrease in the removal of the trajectory plus average pose which drops the recognition accuracy to 45%. Comparing the removal of variations and trajectory in combination with the average pose, show that the general trajectory of the walker contains much identifiable information in their overall characteristic, while the small variations from the trajectories are only meaningful when their dynamic features are preserved.

Lastly, we look at the same feature combinations as before but this time for sex recognition (see Fig. 4.8). In the case of the combination of macro and static features, the removal of the variations does not lead to a drop in accuracy, while both combinations with coarsening macro drop to the same accuracy level of about 90%. This suggests that obfuscations in combination with macro features have a bigger impact on the accuracy in comparison to combinations with static features. All macro-dynamic combinations result in a decrease of performance to about 90%.

Furthermore, removing the variations plus macro coarsening increases the performance slightly when compared to just performing the same macro coarsening alone (see Fig. 4.4). In the micro-static combinations, we find the removal of the trajectory average pose combination to create a large drop in accuracy.

4.3.5. Utility

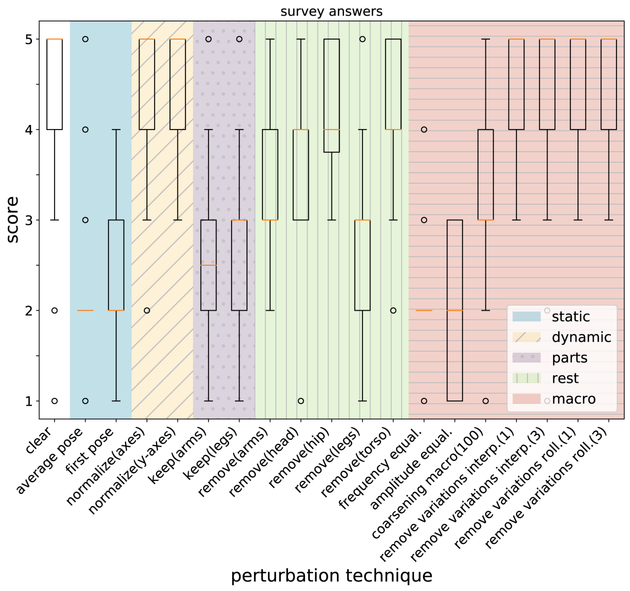


Figure 4.9.: Boxplots of the naturalness rating scores for the perturbation techniques that retained utility.

Finally, we report the results of the naturalness evaluation for all perturbation techniques that have a median rating score which is greater than 1 (all techniques that retain some utility; on the 1-5 scale described in Sect. 4.2.5) and are shown in Fig. 4.9. Perturbations that resulted in a median score below that, were assumed to retain no utility and are therefore not plotted. First, we note that none of the micro feature perturbations retained any gait naturalness. In the static category only average and first pose managed to appear minimally natural, with median naturalness scores of 2. The exclusion of body parts of the walkers had deleterious effects on the perceived naturalness, while still maintaining some level of naturalness depending on the specific removed body part. Interestingly, keeping only the arms or legs of the walker was rated as still somewhat natural, whereas all other individual body parts in isolation were rated as non-natural.

The normalization of all axes and the normalization of the y-axes achieve the same level (median of 5) of naturalness as the clear data. The only other techniques that achieve the same naturalness ratings are the remove variation techniques. In general, these results are within our expectations, as perturbing the data should either maintain the same level of naturalness or decrease it. The fact that most of the macro features retained the naturalness of the walker is also unsurprising, as they preserve the majority of the gait variations while the small variations we kept in the micro features are not perceived as natural anymore. We did not evaluate the naturalness of the combinations, however, we assume that a combination will at most reach the minimum naturalness rating score of its two used perturbation techniques.

4.4. Discussion

Using ML for gait recognition based on motion capture data, we investigated the importance of features based on findings in psychology for identity and sex recognition. The findings reported here, suggest that all of the features reported by psychology are transferable to ML approaches in identification performance based on walking motion. The identification procedure is robust as even when large parts of the data are removed the identification rates are high, only when multiple features are removed from the data a significant impact on the accuracy can be observed. Consistent with previous studies in psychology and neuroscience [279, 395, 361], we found that dynamic and static features contain much identifiable information, hinting at strong temporal and physiological dependencies in the data.

We anticipate that for the development of suitable anonymization techniques for gait data the dependencies between the features have to be accounted for, as otherwise, the reconstruction of the clear data is likely possible. For example, noise applied to marker positions could be removed by smoothing the trajectories over time, since the general gait cycle remains intact. Therefore, when changing the position of markers, the subsequent time steps must also be considered. Alternatively, when removing a marker from a pose, it can be reconstructed from the positions of the remaining markers. Wang et al. [371] have convincingly demonstrated this, showing how adding noise does not effectively perturb correlated data.

Interestingly, the removal of body parts and the subsequent performance accuracy alone indicates a high redundancy in the data, and as such focusing on a single feature for anonymization is unlikely to achieve a meaningful anonymization effect. This effect, albeit in a much weaker form, has previously been shown in human person and biological perception studies: The elimination of some local information, for example by removing *point-light display* (PLD) dots corresponding to body parts, does not affect the recognition as long as a certain degree of global form revealing dynamic posture changes is preserved [27, 176].

Both, the overall trajectories of the gait as well as small variations in the data, allow for recognition of individuals. Thus, making it necessary to adjust the overall gait trajectory for anonymization purposes. The overall pattern of results here provides converging evidence for the need to consider gait motion capture a strong personal identifiable trait, even when recorded at low resolution or low frame rate. Many features, as investigated here — macro, micro, dynamic and static features as well as individual body parts — contain strong identifiable information about both, the identity and the sex of a human walker. With our simple ML-based feature perturbation approach we found that coarsening the marker positions precision, with the respective recognition performances of 45% and 2% for sex and identity exhibited the strongest reduction of classification accuracy while removing dynamic & static features generally only reduced recognition slightly. However, our utility evaluation of the features shows that the perceived naturalness of the perturbed data is diminished when the general motion or body structure of the walkers is removed. Thus we see a strong indication that in order to develop strong anonymization for gait data, while keeping its utility intact, a holistic approach is required. Such an approach should take the dependencies in the data and the requirement for natural-looking results into account, for example by generating synthetic gait trajectories.

4.4.1. Limitations and Future Work

The present study is based on data from 57 young adults. As such, it may be possible to achieve better anonymization results with larger samples, as it becomes difficult to distinguish between individuals. However, as we have shown gait data does contain a large amount of identifiable information, so larger effects from bigger samples are unlikely. The present work presents results on one sole gait cycle per sample, future work should include multiple sequential gait cycles or gait data from multiple sessions. Furthermore, as all individuals were from a similar age cohort, we believe that having a cohort of individuals who are very similar to one another also strengthens the recognition results, as it becomes more difficult to distinguish between them. It is possible that with the improvements of machine learning approaches, better classification results can be achieved on our perturbed data. As such our approach only gives a lower bound how much identifiable information remains in the perturbed data. This fact is also shown by some of the combinations of perturbation techniques where the combinations achieved higher recognition accuracy than the individual techniques alone.

With regards to the user study we would like to point out that our definition of utility only takes into account how natural other people perceived the anonymized PLD gait sequences shown to them. We did not investigate if the original walker themself would find their perturbed gait to be natural. We did so because we assumed that the device used in our system-and-threats-model is trusted by the user and therefore would display the real gait (pre-transfer to the service provider described in Sect. 3.1; labeled "clear" in the present work) to the user as it is recorded locally in real-time, instead of an anonymized version of the user's gait. Furthermore, we based our present investigation on an existing open-source dataset and therefore have no access in an ethical and legal way to the original walkers due to inter alia data protection and privacy reasons. Future studies that obtain their own motion capture recordings could include an evaluation of utility by asking the recorded walkers themselves to evaluate their perturbed gait or other movements recorded with motion capture.

For additional future work we propose to conduct the same set of experiments with human observers to directly compare human and machine gait recognition, in order to gain insight into how both differ in regards to identifying individuals and their sex. Although, the human ability to process biological motion such as gait-based person perception and recognition is susceptible to viewer-specific influences such as age [35], social factors (e.g., interpersonal context, stereotypes) [36, 153], neurodevelopmental disorders (e.g., autism, schizophrenia) [36], and other potential experimental , concomitant, and individual factors[93, 122, 411, 59, 153, 191]. Thus, utilizing machine gait recognition provides a more objective evaluation method for different anonymization techniques.

4.5. Chapter Summary

In this chapter, we addressed the question of how much specific features of human gait contribute to the ability to discern the identity or sex of different human individuals in gait data. Here, we found that overall identification performance was indeed very robust. Removing large parts of the data, either by omitting body parts or reducing spatial and temporal resolution, did have little effect on the recognition performance.

One possible interpretation of the findings is that gait is idiosyncratic and very redundant. Moreover, gait can be considered an individual trait that shows little variability over time and even lifespan. Studies reported that major adult gait emerges already at the age of five years, although age-related effects such as slower gait or shorter steps as well as age-related body proportion changes have been found as well [242].

Our results suggest that gait will be very hard to anonymize effectively. This entails that anonymization cannot be achieved with simple means, but will require intricate approaches that take the inter -dependency of the connected body, as well as the overall generating process of the walking human into consideration. Utility can only be retained when the macro structure of the walker and its dynamic are largely kept intact.

5. Investigating the Privacy Issues of Facial Motion Data

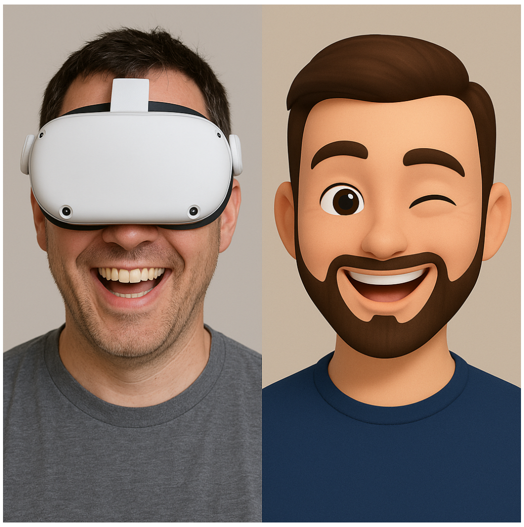


Figure 5.1.: A user wearing a mixed reality headset with facial motion tracking. Their avatar mimics their facial expressions. This image was generated with ChatGPT-4.5.

Mixed Reality (MR) promises to fuse the real and digital worlds. This implies the universal tracking of MR users to create precise digital twins of them. Their appearance, voice, and motions are captured and streamed onto digital avatars. The newest generation of MR headsets (e.g., Apple Vision Pro[1] and Meta Quest Pro[2]) already integrate face and eye tracking to animate the faces of these digital avatars (see Figure 5.1 for an example). In this process, the videos captured by the headsets are transformed into abstract representations of facial motion data known as 'blendshapes'. Integrating facial and eye motions improves social interactions in MR, as subtle non-verbal cues can now be transmitted to a dialogue partner. Currently, we are still in the early adopter stage of this technology as only a handful of applications such as VRChat[3] or virtual YouTubing (using a virtual character

[1] https://www.apple.com/apple-vision-pro/

[2] https://www.meta.com/de/en/quest/quest-pro/

[3] https://hello.vrchat.com/

This chapter is based on the contributions:

- Adriano Castro, **Simon Hanisch**, Matin Fallahi, and Thorsten Strufe. "FacialMotionID: Identifying Users of Mixed Reality Headsets using Abstract Facial Motion Representations". 2025. DOI: 10.48550/arxiv.2507.11138.

to create videos) make use of facial motions. Nonetheless, with the advancement of MR, facial motion tracking is expected to become a standard feature of future MR devices.

However, sharing facial motion data in MR poses a potential privacy risk because facial motions are a behavioral biometric trait. It may yield both identity and attribute disclosure risks: An attacker could use the facial motion data from the avatar shared in MR to perform privacy inferences like identification or employ attribute inferences, like emotion recognition.

Imagine a user visiting a digital store in the Metaverse wearing an MR headset. The user has a generic avatar that does not reveal their identity, and the avatar's facial motion tracking is turned on by default. Without the user's knowledge, the store owner can collect their facial motion data by observing the avatar's facial animations. The store owner can use this data to identify the user, determine if they have visited the store before, and recognize their facial expressions to see which items they like. Thus, the user shares much more private information than they realize.

Although many behavioral biometric traits, such as gait [115], voice [53], and eye gaze [173], are already known to be privacy sensitive, this remains an open question for facial motions. Therefore, we seek to understand whether individuals can be identified from facial motion data and whether emotional states can be inferred.

The main contributions of this chapter are as follows:

- We recorded a novel facial motion dataset, which for the first time allows the investigation of associated privacy risks.
- We demonstrate that the identification of individuals is possible from facial motion data alone.
- We show that re-identifying people across sessions and different MR headsets is possible.
- We confirm that emotion recognition from abstract facial motion data can be performed with high accuracy.

5.1. Related Work

In the following section, we will present the related research on identifying individuals in MR through analysis of facial motion. The primary focus of our research lies in the development and analysis of methodologies for the identification of facial motion from video data, the detection of facial expressions, the identification through eye gaze, and the identification of individuals from MR motion data.

Facial Motion Identification Some preliminary studies have been conducted on the identification of individuals based on facial motion, with the majority of these studies focusing on video data.

Benedikt et al. [29] employed 3D videos of faces to assess the distinctiveness of facial motion for biometric authentication. The trajectory of these facial motions is then represented within the Eigenvector space of diverse facial expressions. Their findings indicate that non-verbal tasks may not be as effective in terms of identification from facial motions as verbal tasks. Zhang et al. [403] performed a similar study, in which they collected 3D videos of participants speaking a passcode 10 times. The system demonstrates an impressive capacity to identify the participant from the dynamic features of the video, achieving a 96% accuracy rate with 77 participants. Haamer et al. [110] collected a video dataset of 61 participants performing various emotion tasks. They then show that participants can be identified using the videos recorded.

Moreira et al. [246] utilized a neuromorphic sensor, an advanced device capable of capturing precise alterations in individual pixels, to record the facial expressions of 40 participants while reciting nursery rhymes. They can show that identification is possible with accuracies as high as 96%.

The existing literature suggests that the identification of individuals through facial motion is feasible for both facial expression and speaking tasks. However, given that the majority of studies employ video data, it remains uncertain whether identification can be achieved exclusively through the analysis of facial movements alone, since face recognition is possible on static face images. Additionally, the question remains open whether individuals can be identified across multiple sessions via facial motion data.

Facial Expression Recognition One field of study that has focused on facial motion analysis is facial expression recognition. The objective of facial expression recognition is to categorize the emotions displayed by the individual captured on video [169]. Zhao et al. [406] propose a lightweight model to extract the displayed emotion from face images. Wen et al. [374] use an attention network to perform emotion recognition and achieve state-of-the-art performance. Furthermore, Chen et al. [51] have employed the differences between a neutral face and an expressive face to enhance the learning of different face expressions. To improve generalization in their face recognition model, Zhang et al. [405] propose learning an identity-independent representation of facial expressions using deviation learning. This involves subtracting a person's identity, established by a face recognition model, from their facial expression embedding.

Lee et al. [179] investigate facial expression recognition using a face mask that measures facial deformation, rather than via videos.

Facial expression recognition has also already been investigated in the context of MR by Chen et al. [50] in a study in which extra cameras have been integrated into an existing MR headset. Additionally they used an external camera to capture the part of the face which is not hidden behind the MR headset. They then show that they can achieve a facial expression recognition accuracy of 95%.

Facial expression recognition shows that facial motion data is useful for more than just direct social interactions between people. However, this information should also be considered private, and individuals should have the choice of when and how they share their emotions.

Eye Gaze Eye gaze was recognized as a privacy-sensitive topic some time ago and has also drawn attention as a possible behavioral biometric trait for authentication. Lohr et al. [193] showed that they could identify 269 subjects with a mean EER of 4.72% using the SBA-ST dataset [94], which was captured with a dedicated eye tracker. They further improved their method in EyeKnowYouToo [194], which is the current state-of-the-art model for user authentication based on eye gaze. They achieved an EER of 3.66% at a sampling rate of 1000 Hz and an EER of 8.77% at a sampling rate of 125 Hz. In a later study, Raju et al. [302] investigated the performance of eye gaze authentication on the GazeBasedVR [192] dataset and showed that short-time authentication works well but that the EER increases to 10% for longer sessions.

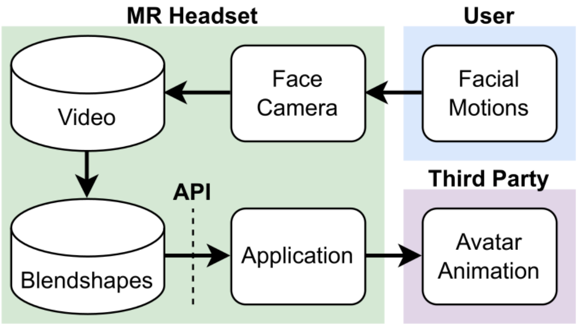


Figure 5.2.: The data sharing pipeline of facial motion data captured by MR headsets.

Shao et al. [328] aim to create an eye-gaze identification system in MR that is independent of the content shown to users. They use two encoders: one for content and one for eye gaze. They achieved an F-score of 92%. Asish et al. [20] use eye gaze features of 34 people performing four different tasks for identification in *Virtual Reality* (VR).

As the privacy-sensitive nature of eye gaze data has been recognized, the first studies [187, 307, 375] seeking to anonymize it have emerged. Common methods of anonymization include adding noise or smoothing the eye gaze trajectories.

Eye gaze is useful not only for authentication, but also for foveal rendering. Foveal rendering is a selective rendering process that increases the level of detail in the section of the image at which the user is looking. Several studies [18, 138, 75] attempt to predict eye gaze to enable foveal rendering.

The research on eye gaze data showcases the dual nature of behavioral biometric data, as both privacy inferences, as well as desired applications like authentication and foveal rendering are possible with it.

Mixed Reality Identification In recent years, the subject of identifying people using motion data recorded by MR headsets has gained traction, and multiple studies have been published on the topic. Among the first of these studies, Miller et al. [240] recorded 511 participants watching 360-degree videos in VR. The researchers demonstrated a high identification rate of 95% using the head and controller motions. Liebers et al. [184] demonstrated that identifying individuals is possible by combining the head orientation and eye gaze of 12 people captured with a MR headset. Moore et al. [245] investigated which VR tasks are most effective for identification, once again using headset and controller motions. They found that identification success depends on the VR content used. Nair et al. [250] used a large-scale dataset of people playing Beat Saber[4] and demonstrated their ability to identify players in a pool of over 50,000 people with 94% accuracy using 100 seconds of headset and controller motion data.

5.2. Background

Here, we briefly describe the background for MR motion tracking required for this work.

Facial Motion Tracking: The MR headsets used in our study rely on camera-based face tracking. Inward-facing infrared cameras capture the eyes and mouth of the person wearing the headset. This video data is then transformed into a symbolic representation

[4] https://www.beatsaber.com/

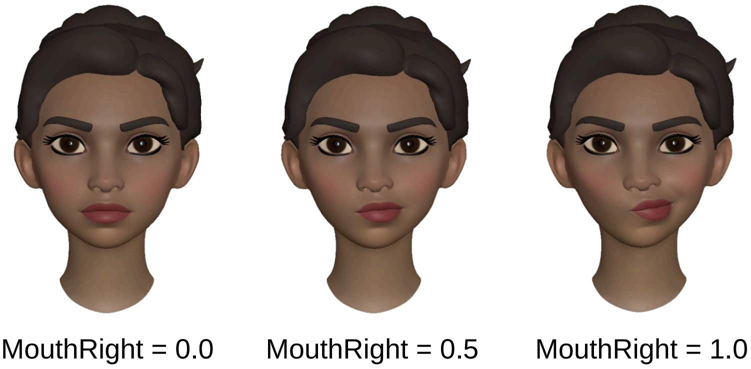


Figure 5.3.: A blendshape named "MouthRight" being activated on an MR avatar from 0 to 1 through interpolation.

which is shared via applications on the MR headset. See Figure 5.2 for the full data sharing pipeline of facial motion data. For facial motions, the data is represented as blendshapes. Blendshapes are a type of interpolated animation, also known as morph target animation. In this type of animation, the neutral state and deformed version of an object are stored for each blendshape. Then, for each frame of the animation, the object's vertices are interpolated between the neutral and deformed versions. An example of a blendshape for an MR avatar is the right part of the mouth (see Figure 5.3). In the neutral state, the mouth is symmetrical; in the deformed state, it is pulled to the right side of the face. All intermediate states can be created via interpolation. The blendshapes defined by the MR headsets are usually based on the *Facial Action Coding System* (FACS) [83, 132]. The former is a system that defines and describes all distinguishable facial movements, so-called action units. These action units are derived from anatomy, and with them complete expressions can be recognized objectively. Two examples of such action units are "Cheek Raiser" and "Lip Corner Puller" that together can be interpreted as the expression of happiness.

Eye Tracking: The user's eye gaze is captured via infrared cameras positioned inside the MR headset. The video is then converted into gaze direction and eye position data.

Motion Tracking: For the motion tracking in MR headsets there exist two main approaches. Inside-out tracking describes the approach in which multiple cameras on the outside of the headset are used to establish its position. The second approach is light house tracking in which one or multiple static light houses emit sequences of infrared light which are registered by infrared sensors on the surface of the headsets and the controllers. The headset and controllers can then compute their distance and orientation in relation to the light houses. When comparing the two approaches, inside-out tracking is less precise but easier to use than lighthouse tracking.

5.3. Study Design

In this section, we describe the design of our study to investigate identity and attribute disclosures from abstract facial motion data. We first explain the general rationale before providing a more detailed explanation of the tasks used and the selected recording schedule.

Design Rationale The main goal of our study is to investigate whether identifying individuals from their facial motion data is possible. To allow biometric recognition systems to train on the data and recognize identifying patterns, we require a large number of samples. Therefore, we require numerous repetitions and task executions involving a diverse group of participants. Additionally, we aim to determine whether facial motion data is a stable biometric factor over time; therefore, we will record multiple sessions with each participant. Lastly, we want to investigate whether facial motion data generalizes well when different devices are used to capture facial expressions. Therefore, we record our participants using multiple device types that integrate facial motion tracking.

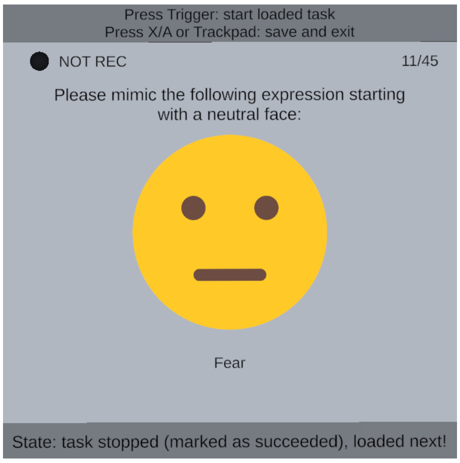
We see the main application of facial motion data for animating digital avatars as speaking to other people and displaying emotions. Consequently, we focus on tasks involving two types of categories, namely speech and emotional expression for data collection. As mentioned in Section 5.1, emotion recognition has been shown to work previously. Hence, we integrate it into the study to test collected data and to compare results. Since facial motion data will likely be used in combination with eye gaze and head motion data—and as these are readily available in the common MR headsets—we also collect these.

Recording Procedure We chose to record our participants over the course of three separate sessions, with each session being approximately a week apart from each other. In the first session, participants first answer a short questionnaire about demographics before the actual recording starts. During each session, we record each participant performing the same set of tasks with two different MR headset types. We chose to keep one headset type the same throughout all sessions, whereas the respective other headset was alternated between the remaining two in each session. This allowed us to record all participants using three different headsets. Due to the change in the second headset, we split our participants into two groups, A and B, to keep track of which second headset had to be used in each session.

Tasks We designed a task-based study in which participants performed predefined tasks sequentially. An overview can be seen in Table B.1. At the beginning of the study, one tutorial task was performed for each task type. To cover the described applications, we selected verbal tasks, in which participants read a given text aloud, and non-verbal tasks, in which participants mimic a facial expression. Studies such as [29, 246] have demonstrated that verbal tasks contain the most identity cues in facial motion, unlike non-verbal tasks. Therefore, the predominant task category we selected is verbal tasks.

First, the participant is shown the current task. Then, the participant starts the actual recording phase for the task by pressing a button. During the recording phase, the participant performs the task. The recording phase is ended by pressing the same button again. All tasks and their repetitions are presented to the participant in a random order. There are four repetitions for each task in the first session and five repetitions for each task in the second and third sessions. The reduction of repetitions in the first session allows time for the questionnaire.

Non-verbal Tasks: We presented the non-verbal tasks using emoticons that displaying three different facial expressions: happiness, anger, and fear. See Figure 5.4+5.5 as an example for a non-verbal task. This abstract representation should encourage participants to perform the facial expressions as they normally would rather than closely mimicking



Press Trigger: start loaded task
Press X/A or Trackpad: save and exit

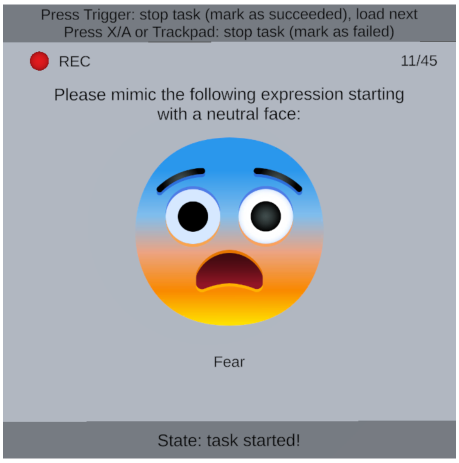
● NOT REC 11/45

Please mimic the following expression starting with a neutral face:

Fear

State: task stopped (marked as succeeded), loaded next!

Figure 5.4.: The participant performs the expression starting with a neutral face.



Press Trigger: stop task (mark as succeeded), load next
Press X/A or Trackpad: stop task (mark as failed)

● REC 11/45

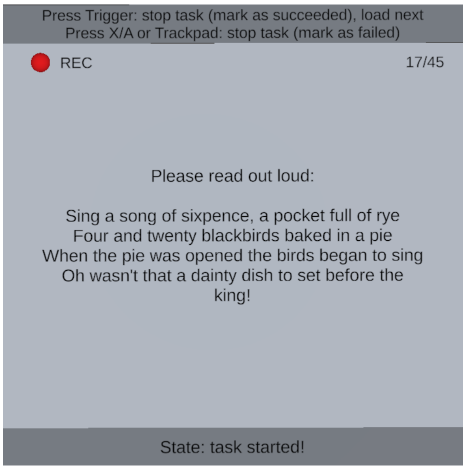
Please mimic the following expression starting with a neutral face:

Fear

State: task started!

Figure 5.5.: The participant performs the expression fear, after doing the neutral face .

the avatars shown to them. Therefore, we did not use high-fidelity digital avatars. We instructed participants to mimic the non-verbal tasks shown to them by starting with a neural facial expression and to then transition into the shown facial expression. An animation of the emoticon changing from neutral to the target expression illustrates this process.



Press Trigger: stop task (mark as succeeded), load next
Press X/A or Trackpad: stop task (mark as failed)

● REC 17/45

Please read out loud:

Sing a song of sixpence, a pocket full of rye
Four and twenty blackbirds baked in a pie
When the pie was opened the birds began to sing
Oh wasn't that a dainty dish to set before the
king!

State: task started!

Figure 5.6.: An example of a verbal task in which the participant is uttering the nursery rhyme "Sing a Song of Sixpence".

Verbal Tasks: During the verbal tasks (see Figure 5.6), participants are asked to utter words and sentences. Lu et al. [201] have shown that words and groups of sentences that contain a large number of phonemes are best suited for identification. A phoneme is the smallest unit of sound which makes a lexical difference in a language. Additionally, Moreira et al. [246] have already shown that reciting nursery rhymes are suitable for facial motion identification. Therefore, we selected nursery rhymes for the verbal tasks because they contain various repetitive phonemes. To select the nursery rhymes, we used a list[5]

[5] https://www.bbc.co.uk/teach/school-radio/articles/z4ddgwx

of common English nursery rhymes. To keep the verbal task short, we prepared the list by splitting all rhymes, such that each part is at most four lines long. Next, we counted the phonemes of each nursery rhymes and selected the top three with the highest count. Out of these selected nursery rhymes, we selected one word of each that contained the highest amount of phonemes, constituting the word tasks.

5.4. Study Implementation

The study was conducted between January 22 and February 14, 2025. It took place in the kd2lab that contains multiple small booths specifically designed for user studies, and an office for their supervision. We divided each study day into 12 slots, with each day ranging from 8:30 am to 6:15 pm. Since we aimed for a study duration of approximately 30 minutes, an equal time allocation was assigned to each slot. To compensate for unexpected duration times, we added a 15-minute break between each slot. At each slot, two individuals participated simultaneously — one from group A and another from group B. As each booth contained a door, each participant could perform the study without any disturbances.

Ethics The data collection was approved by the ethics commission of the Karlsruhe Institute of Technology (research project ”Privacy of Facial Motions”) and was conducted in accordance with the Declaration of Helsinki. Participants were paid based on their time of participation at an hourly rate of 14€. Additionally, participants received a flat bonus of 2€ or 3€ for participating in the second and third sessions, respectively. We obtained informed consent from all participants for the data collection and processing.

Apparatus During the study, we used four MR devices, namely two Meta Quest Pros, one Pico 4 Enterprise[6], and one HTC Vive Pro Eye[7] with the Facial Tracker add-on[8]. All of these devices support eye and facial tracking in addition to standard head and controller tracking. Moreover, the devices and their tracking are supported by Unity, the Game Engine that we used to implement the application for our study. While the first device type is designed for both augmented and virtual reality, the other two are purely VR devices. Since we only require VR, the three types of devices were deemed suitable for our experiments.

The study was implemented as a Unity application since all selected MR devices supported it. Unity Engine v2021.3.32f1 was utilized for development, as it was the most recent long-term support version supported by all headsets and their tracking APIs. We created a scene for each device, as they required individually configured XR cameras and device specific code to activate their motion tracking.

To be able to access the motion data of the devices and store them, we utilized several Unity packages that allowed the interaction with the **APIs** of the devices. For the Meta Quest Pro we used the Meta Movement SDK v71.0.1 including the Meta XR Core v71.0.0 and the Meta XR Interaction SDKs v71.0.0[9]. For the Pico 4 Enterprise we used the PICO Unity Integration SDK v2.5.0[10]. And for the HTC Vive Pro Eye we used the VIVE OpenXR Plugin v2.0.0[11] with addition of the VIVE SRanipalRuntime v1.3.1.1 and

[6] https://www.picoxr.com/global/products/pico4e

[7] https://www.vive.com/sea/product/vive-pro-eye/overview/

[8] https://developer.vive.com/us/hardware/facial-tracker/

[9] https://developers.meta.com/horizon/documentation/unity/move-overview/

[10] https://developer.picoxr.com/document/unity/?v=2.5.0

[11] https://github.com/ViveSoftware/VIVE-OpenXR-Unity

the OpenXR Plugin v1.9.1 for the facial tracker. Both our Meta Quest Pros used during the study had identical software and runtime as well as OS versions, namely v71.0.0 and SQ3A.220605.009.A1 respectively. The Pico 4 Enterprise ran on version v5.9.9, and the Vive's eye and lip camera versions were v2.41.0-942.e3e4 and v50100 in corresponding order.

Recruitment We recruited 116 participants (45 female, 71 male; age mean 23.6 years, std 4) with the help of the KD2Lab panel of the Karlsruhe Institute of Technology. The distance between two subsequence sessions was between 4-16 days (participants per session 1: 116, session 2: 83, session 3: 49). Of the participants, 67 were native German speakers, while the rest reported a different mother tongue. 67 describe themselves as ambiverts, 26 as extroverts, and the remaining 23 as introverts.

The participants were assigned to their respective group at random. While group A used the HTC Vive Pro Eye in addition to their assigned Meta Quest Pro in the first session, group B started with the Pico 4 Enterprise. In the second session, group A then received the Pico 4 Enterprise instead of the HTC Vive Pro Eye, and group B vice versa. In the third session, group A and B each returned to their first headsets. Thus, each participant who participated in all sessions used each device at least once and the Meta Quest Pro three times.

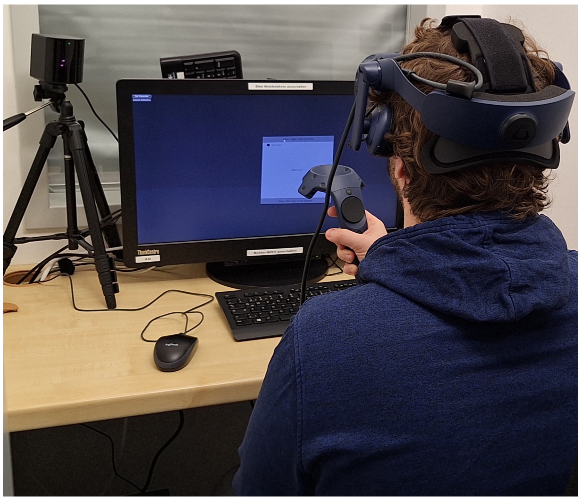


Figure 5.7.: A participant performing the tasks with the HTC Vive Pro Eye.

Session Procedure For the first session, our participants required more thorough guidance and support. We began by introducing the study and explaining the procedure, emphasizing the data collection process and its purpose. Then, we started a timer to keep track of their study duration, which was relevant for their payment at the end. Then, we assigned each participant a random pseudonym to be used for the remainder of the study.

Next, we escorted each participant to their assigned room. Each participant was given an information sheet with details about the study, a data protection agreement, and a survey. The survey collected information about the participants' age, sex, origin, self-assessed personality traits, English proficiency, and mother tongue. After completing the survey, the participants watched a short introductory video showing them how to use the MR headsets and their respective calibration procedures.

After watching the tutorial videos, the participants were brought to the booth with their first headset. We helped them become accustomed to the headset and to perform the eye calibration. Thereafter, the participants started the Unity application and, thus, performed the tasks shown through their MR headset. When they completed the tasks with the first headset, they were brought to the second one, where we repeated the procedure. At the end, the participants filled in a short online survey to receive their payment with their own payout token assigned through the experiment organization. To reduce any possible bias in the data due to headset order, the order of the headsets was inverted for each group of participants. See Figure 5.7 for an example how the participants performed the study.

In subsequent sessions, participants did not have to fill out the survey or data protection sheet again. Although we asked the participants if they wanted to watch the eye calibration tutorial videos again, they usually skipped them since they remembered how to perform the tasks. Additionally, the subjects usually skipped reading the study information sheet from the first session. They were usually brought directly to the headsets and performed the study as described above.

Troubleshooting During the study, there were some difficulties. For the first recording day (22.01.2025) we encountered a problem for the facial motion recording of the HTC Vive, and as a consequence the HTC Vive recordings for the first day contain less blendshapes then the following recordings. Another problem we encountered with the HTC Vive was that for some of the audio recordings the recording frequency was higher than configured, though this was unproblematic since our data processing approach presented in Section 5.5 is robust against it. The eye calibration of the Meta Quest Pro devices turned out to be challenging, as it would regularly finish unsuccessfully. This seemed to be more frequent with participants wearing glasses, yet it also happened with non-glasses wearers. In such problematic cases, we helped the participants adjust the lenses and the position of the headset on their heads — it did help a relative number of cases, but not all of them. Due to these problems, the quality of the eye tracking for the Meta Quest Pro suffered. Another challenge was that both the eye and face tracking of the Meta Quest Pro devices tended to suddenly malfunction in between participants. This happened once per Meta Quest Pro device, and was unfortunately only discovered at the end of the day. Due to this issue, we lost 19 recordings.

5.5. Data Processing

Upon the completion of each participant's session, our Unity project generated a unique directory containing the relevant data and metadata gathered during it. This included the unsegmented face, eye, and head motion data, as well as the execution order and timestamp range of each task repetition, a microphone recording along with its metadata, and a log file.

Since the facial and eye motion data formats exported by the MR devices are not exactly the same, a unification step was necessary. See Table B.2 for the exact mapping for each MR headset. For n-to-1 mappings from the devices to the unified format, we use the mean of the directions. One example of this is the CheekPuff blendshape. The HTC Vive and Meta Quest Pro support CheekPuff for both sides of the face, while the Pico 4 Enterprise only returns one CheekPuff blendshape.

Table 5.1.: Overview of the dataset regarding the amount of samples it comprises.

Segmentation	Total	Per Group		Per Device			Per Session		
		A	B	Vive	Pico	Meta	0	1	2
Recordings	499	232	267	132	127	240	229	150	120
Tasks	19296	8883	10413	5175	4905	9216	8136	6750	4410
Words	197255	88477	108778	45087	51631	100537	82814	67362	47079

As our study consisted of tasks, we partitioned the unsegmented data of each participant into individual task-level segments. Moreover, we segmented the aforementioned task-level segments which belonged to text tasks further into word- and phoneme-level segments. The word- and phoneme-level segmentation can be found in the Appendix B.2.2.

Task-Level Segmentation: First, the data was segmented by task. To achieve this, we used the timestamp ranges stored during each task repetition. When a participant started a task, a timestamp was saved to mark the start of execution. Then, when the participant finished the task, a second timestamp was saved to mark the end. Since we stored the timestamp of when each sample of motion data was collected, we could identify which samples belonged to which task repetition in each motion data file.

Data Availability In total, we recorded 259 sessions. 19 of these sessions were missing one headset recording, resulting in a total of 499 individual headset recordings. Table 5.1 provides an overview of the number of samples segmented as described above.

5.6. Evaluation

Here, we present the evaluation that we performed on the dataset. Our main goal is to investigate the types of privacy inferences that can be made from facial motion data. However, we also perform the same experiment on eye gaze and head motion data to allow for comparison. First, we present the experiments we performed. Next, we detail the methodology for the biometric recognition system. Lastly, we present the results of the experiments.

Experiments First, we want to establish whether identification from facial motion data collected using MR headsets is possible. Prior work on facial motion videos (see Section 5.1) and on eye gaze identification (see Section 5.1) has demonstrated the feasibility of identifying individuals. Therefore, we expect identification from facial motion data to be possible. For Experiment **E1**, we will investigate the identification for each headset separately, as well as all headsets together. Next, we want to know if the identification is stable over time. For Experiment **E2**, we use the first two sessions for training the biometric recognition system and then only test on the third session. Then, in Experiment **E3**, we examine whether we can re-identify individuals when they start using different headsets. This gives us insight into how dependent identification is on headset type, and whether it can be generalized across MR headset types.

Besides identification, the related work (see Section 5.1) suggests that it should be possible to infer the facial expression and therefore we expect that it is possible to infer the emotion displayed in the non-verbal tasks. In Experiment **E4**, we test how good we can recognize the emotions displayed in our non-verbal tasks. Further, we also test if we can correctly classify which verbal task was performed. More experiments and their results can be found in the Appendix B.2.3.

Data Preparation & Splitting For our evaluation, we use task-level segmentation of our dataset in the unified data format. We filter out the recordings performed on January 22, 2025, as some of the Vive's facial motion data values are missing. We then remove the timestamp column from the remaining samples and resample each one to 100 frames, normalizing the size of all samples.

Next, we split the data into training and testing datasets for the biometric recognition model. The testing dataset is used exclusively to calculate the model's final performance. Since different experiments require different data splits, we use multiple splits:

Random: For the random split type, we randomly split all samples, allocating 80% to the training dataset and 20% to the testing dataset.

Session: For the sessions split-type, we use the recordings from the first two sessions from each participant as the training dataset and the last session as the testing dataset.

Leave-one-headset-out-per-participant (LHPP): The LHPP split type uses two MR headsets per participant for the training dataset and one MR headset for the testing dataset. This allows the biometric recognition model to learn to recognize specific participants and to use data from each MR headset type.

Participant: The participant split type allocates 80% of participants to the training dataset and 20% to the test dataset. This type of split is used for attribute inference experiments to prevent cross-contamination of the results, e.g. the model learning to recognize attributes by identifying the specific participant.

Biometric Recognition Models As identification using facial motion data is a new field, it is unclear which machine learning approach will perform best for the biometric system. For our experiments, we therefore use three commonly employed machine learning models as a biometric recognition system. The first is a **simple** fully connected neural network that receives each sample as a single vector. This neural network consists of at least two fully connected linear layers and a variable number of hidden layers, which are determined via hyper parameter optimization. After each linear layer, we use a *Rectified Linear Unit* (ReLu) activation function, as well as a dropout layer, to prevent overfitting. The second model is a *Long Short-Term Memory* (**LSTM**) that processes each sample frame-by-frame. To determine the most likely class, we first use a linear layer to reduce the size of the output vector to the number of classes. The third model is **EKYT** [194], a DenseNet-based architecture. Between each convolution block, the network uses batch normalization and the ReLu activation function. All networks use log softmax to perform the final classification step.

The training dataset is randomly split into a main training dataset and a validation dataset for model training. The main training dataset contains 90% of the data, and the validation dataset contains 10%. Each model is trained for a maximum of 100 epochs with early stopping if the validation accuracy does not increase for 10 epochs. We use negative log likelihood loss as the loss function and 1280 samples as the batch size.

We determine the best model parameters for each experiment by performing parameter optimization for 100 steps. See Table B.3 for the optimized parameters. After optimization, we use the model with the best performance on the validation dataset and run it on the testing dataset to determine the final accuracy for each experiment.

Table 5.2.: Identification accuracy using a random split

Data Type	Model	Vive	Pico	Meta	All	Chance
Facial	Simple	0.78	0.83	0.63	0.68	0.02
	LSTM	0.69	0.77	0.59	0.58	0.02
	EKYT	0.94	0.98	0.88	0.9	0.02
Eye	Simple	0.87	0.7	0.45	0.54	0.02
	LSTM	0.86	0.59	0.47	0.49	0.02
	EKYT	1.0	0.87	0.92	0.78	0.02
Head	Simple	0.99	0.88	0.76	0.7	0.02
	LSTM	0.94	0.78	0.76	0.8	0.02
	EKYT	1.0	0.95	0.98	0.95	0.02

Table 5.3.: Identification accuracy using a session split

Data Type	Model	Vive	Pico	Meta	All	Chance
Facial	Simple	0.11	0.26	0.29	0.2	0.04
	LSTM	0.11	0.11	0.24	0.17	0.04
	EKYT	0.14	0.23	0.43	0.27	0.04
Eye	Simple	0.14	0.03	0.03	0.06	0.04
	LSTM	0.12	0.02	0.07	0.05	0.04
	EKYT	0.1	0.06	0.1	0.09	0.04
Head	Simple	0.0	0.02	0.16	0.08	0.04
	LSTM	0.0	0.01	0.15	0.06	0.04
	EKYT	0.0	0.01	0.16	0.07	0.04

Implementation We implemented the biometric recognition models using Python (3.12) and PyTorch (2.6.0). As learning optimizer, we used Adam, and for the parameter optimization we used Optuna (4.3).

Results Here, we present our evaluation results. As a metric, we always use the accuracy, which is defined as the correct classifications divided by all classifications. Further, we also always give the percentage of the largest class in the experiment-specific data split as the chance level.

For our Experiment **E1** (see Table 5.2), we used the random split to gain general understanding of how well the identification works. For the facial data, we find that the Pico achieves the highest identification of 98%, the Vive achieves 94%, the Meta achieves 88%, and using all headsets together we achieve 90%. This fulfills our expectation that identification on facial motion data is possible.

Comparing to the eye and head data types, we find that for both we achieve 100% identification for the Vive and the EKYT model. In general, we can observe that the identification works for all data types, and all headsets, with the head motion data performing the best in general, though the facial and eye motions are not far behind.

Moving on to Experiment **E2** (see Table 5.3), we now split the data according to their sessions into training and testing dataset. In general, we can see for the face data that all

Table 5.4.: Identification accuracy using the LHPP split for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Facial	0.61	0.52	0.63	0.02
Eye	0.45	0.47	0.65	0.02
Head	0.49	0.46	0.63	0.02

Table 5.5.: Emotion recognition accuracy using a participant-wise split for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Facial	0.86	0.86	0.86	0.33
Eye	0.45	0.33	0.59	0.33
Head	0.49	0.32	0.58	0.33

headsets and model combinations exceed the chance level for identification. The best result is 43% balanced accuracy for the face data of the Meta when using the EKYT model. We conclude that the identification across sessions is possible, but most of the learned features from E1 identify the specific session and are not general for the individual. Comparing E1 and E2 results, it is also interesting to see that in E2 the Meta performs far better for facial motions, while in E1 it has the worst performance of all three headset types.

Next, we test if we can recognize participants across different MR headsets in Experiment **E3** (see Table 5.4). The LHPP split leaves for every participant one headset type for which the model has not seen any data, hence, we simulate that the user switches to a new type of MR headset. The best accuracy for facial motion data is achieved by EKYT with 63%, showing that identifying individuals across headsets is possible, however, at a lower rate than in our baseline E1.

In our Experiment **E4**, we tested emotion recognition using only emotion tasks (see Table 5.5). As expected, facial motion data was the most effective for emotion recognition, with 86% accuracy. However, eye and head motions also enabled some emotion recognition, with accuracy rates of 59% and 58%, respectively.

Summary of Results

- We are able to show that persons can be identified from their facial motions.
- The identification across different sessions is possible, however, the achieved accuracy is not on a level that is usable for any real-world system at the moment.
- We are able to show identification across different MR headset types.
- We are able to infer the displayed emotion.

5.7. Discussion

Facial motion data is a behavioral biometric factor that can be used for identification, so it should be treated as such when sharing it online. However, our results indicate that facial motion might not be stable enough to reliable identify individuals over long periods of time. Only larger studies with longer intervals between sessions can determine whether facial motion data poses a long-term privacy threat to individuals. We expect MR headsets to improve their ability to record facial motion data in the future, so we also expect privacy problems with facial motion data to increase.

When we compare our eye gaze and head motion results to those of previous studies, such as GazebaseVR [192] for eye gaze data and Nair et al. [250] for VR data, we find that our identification results are are not as good, especially when considering multiple sessions. We believe this is because the tasks in our dataset are designed primarily to capture facial motion data. For example, GazebaseVR uses specific eye-tracking tasks, such as following a dot with one's eyes or reading tasks. In contrast, we only record data after participants read the tasks and push the button to start recording; therefore, we do not expect much eye motion during recording. Additionally, none of our tasks require head motion, so little variance is expected.

5.8. Chapter Conclusion

In this chapter, we present the results of the first comprehensive study into identifying individuals via abstract facial motion data. We demonstrate that identification rates of up to 90% are possible in a single session. In the multi-session scenario, some identification clues remain, but it is not possible to reliably identify individuals. Our results demonstrate that facial motion data is privacy-sensitive and must be protected accordingly. We expect the collected dataset to be a valuable resource for future research into privacy protections for facial motion data.

6. Improving the Evaluation Methodology for the Anonymizations of Biometric Data

As we found in our survey of behavioral biometric data anonymization (see Chapter 3), the evaluation methodology is flawed. It often implicitly uses an evaluation scenario that does not assume a worst-case scenario, which leads to an overestimation of the anonymization performance. Here, we address this problem.

Reliable evaluation begins at the assumptions made about an attacker. These assumptions must be robust, because otherwise the evaluation methodology will likely deliver grossly inaccurate estimates of anonymization performance. The result will be a false sense of privacy, and the consequence will be the erosion of user trust when anonymizations, perceived as reliable, do not deliver the expected protection. Moreover, any inaccuracy or even error in an evaluation methodology will detrimentally affect advances in research. Flaws in the methodology may feed into future research and thus hinder or even arrest the development of advanced anonymization techniques. The upshot is this: Only when a biometric anonymization technique has been convincingly evaluated can researchers improve on existing techniques or provide privacy-preserving applications to users.

In this chapter, we assess the state-of-the-art evaluation methodology for the anonymization of biometric data. In particular, we assess the evaluative methods for face anonymization and gait anonymization. Our choice for face anonymization is based on the fact that there are many widely-employed techniques in application. We acknowledge that the comprehensive evaluation of any anonymization technique is only possible when utility is also taken into consideration. However, for this chapter, we have narrowed our scope to the improvement of the methods evaluating privacy protection of anonymization only.

Our assessment of the state-of-the-art in evaluation for the anonymization of biometric data shows that these methods often fail at convincingly evaluating the performance of the privacy protection.

The state-of-the-art methods have been uncritically adopted from the evaluation methodology for biometric recognition. In biometric recognition, the problems employ many identities and difficult biometric samples (e.g., profile photos or nearly indistinguishable identities). In anonymization, on the other hand, a difficult problem has a small number of identities which are very diverse, thus making the identities easier to differentiate but more difficult to anonymize.

This chapter is based on the contribution:

- **Simon Hanisch**, Julian Todt, Jose Patino, Nicholas W. D. Evans, and Thorsten Strufe. “A False Sense of Privacy: Towards a Reliable Evaluation Methodology for the Anonymization of Biometric Data”. In: Proceedings on Privacy Enhancing Technologies. 2024. DOI: 10.56553/popets-2024-0008.

Furthermore, the state-of-the-art methods rely on weak adversary models. These methods assume that the attacker is unaware of the anonymization mechanisms in place. For example, a method will use pre-trained recognition models which perform well on clear data. However, such models prove incapable of adapting to data modifications performed by an anonymization technique. Consider this straightforward scenario: An anonymization technique for a face image performs consistently the same block permutation. This anonymization can easily be removed with the inverse permutation. However, the permutation will go unnoticed by a recognition model pre-trained on the clear data. Moreover, if only a single recognition is used, then that will jeopardize the reliability of the evaluation. Although a given anonymization technique may successfully degrade the performance of one recognition system, *other* systems classifying *other* feature vectors may be more robust or even largely unaffected. However, the use of just a single recognition system is the norm among state-of-the-art evaluation methods.

The contributions in this chapter are as follows:

- We assess the current state-of-the-art evaluation methodology for biometric data anonymization and point to fatal flaws in the evaluation methodology.

- We update the state-of-the-art evaluation methodology. Our methodological improvements involve (1) retraining the recognition system on anonymized data, (2) using multiple recognition systems to evaluate the anonymization, and (3) generating evaluation datasets that are challenging to anonymize and consequently reliable for the evaluation of the anonymization performance.

- We test our methodological improvements on the biometric traits face and gait with extensive experimentation. Our evidence supports the conclusion that our improved methodology delivers reliable evaluations of biometric data anonymization.

6.1. Related Work

Biometric recognition spans dozens of biometric traits and hundreds of techniques, but the methodology for evaluating the performance of these techniques has been assessed by only a very few works. First, we will examine the proposed improvements to the stylometry evaluation methodology. Next, we discuss works on the evaluation of face and voice anonymization, which are more closely related to our goals.

Goga et al. [102] assess the methodology for evaluating matching techniques of profiles from different social media platforms. They find that evaluation commonly overestimates the performance of the approaches by using an unrealistic methodology. Granger and Gorodnischy [106] describe the methodology that should be applied to evaluate the performance of biometric recognition for video surveillance applications. For the evaluation of stylometric authorship attribution, Stolerman et al. [343] make the case that an open-set model should be applied since in a realistic scenario the actual author might not be on the suspect list. Brennan et al. [42] propose adding attacks to the methodology of stylometry evaluation because most methods cannot defend against attacks. These investigations of the evaluation methodology in different fields have shown that wrong assumptions lead to an overestimation of performance. In the case of anonymization, overestimation of performance may give users false assurances of privacy because, in fact, their identities are actually left unprotected. In this chapter we similarly look at a current evaluation methodology, highlight issues and propose solutions.

Le et al. [178] discuss how to evaluate privacy-utility trade-offs for face anonymization, but their focus is exclusively on measuring the utility and not privacy.

Recent works [348, 117, 355] propose attacks on biometric data anonymization that use machine learning to reverse the obfuscation of images. These results show that the method is highly effective even when a human observer cannot recognize anything at all in the image. The reversal of anonymization is indeed comparable to the training of a recognition system on anonymized data. However, we consider training recognition systems on anonymized data the more straightforward way to test whether identifying information remains in the anonymized data. Further, we also consider the reduction of the dataset.

In the context of the VoicePrivacy challenge [356], other recent works have investigated the evaluation methodology of speaker anonymization. Noé et al. [261] also propose a framework to evaluate and compare speech pseudonymization approaches using ZEBRA [254] and voice similarity matrices [260]. ZEBRA aims at creating a worst-case metric to evaluate speaker anonymization and voice similarity matrices allow to compare how well specific identities are anonymized. Bonastre et al. [38] propose a benchmarking methodology to test speaker recognition against spoofing and anonymization. We investigate whether some of the methodological improvements to the evaluation of speaker anonymizations, like training recognition systems with anonymized data, can be applied to a wider range of biometrics like face and gait data.

In sum, many improvements to the evaluation methodologies of different research fields have been proposed. However, for the anonymization of biometric data, we find that multiple improvements can still be made to evaluation methodology, such as anonymized data in the training dataset and a more challenging anonymization scenario.

6.2. Improving the Evaluation Methodology

In this section, we aim to achieve a reliable evaluation methodology for the anonymization of biometric data. Our premise is that an evaluation methodology for anonymization techniques should be pessimistic and assume a strong adversary based on the worst-case performance of the anonymization technique.

6.2.1. State-of-the-Art Evaluation Methods for the Anonymization of Biometric Data

We began by gaining an overview of the problems of the state-of-the-art evaluation methods. To this end, we assessed the papers covered in our survey (see Chapter 3) plus the papers of another survey by Ribaric et al. [311] on the topic of biometric data anonymizations. Next, to gain a closer perspective on the field of face anonymization, we analyzed works published from 2018 [301, 88, 152, 370, 140, 223, 327], and one work from 2005 [259]. We included as many works as we could find which appeared at *USENIX Security*, *Privacy Enhancing Technologies Symposium* (PETs), and *Data and Applications Security and Privacy*. As a recent work [177] of 2023 testifies to the persistence of the said methodological flaws to this day.

Our survey shows that the methods for evaluating techniques of biometric recognition or anonymization use the same recognition systems, the same datasets, and the same evaluation scenarios. This unquestioned reuse of the same attacker model, dataset, and scenario is highly problematic and will undermine the reliability of any evaluation of anonymization performance. Our reasoning is as follows. In biometric recognition, an evaluation method presents challenging scenarios to the recognition system. Identities are hard to distinguish from one another, the number of identities to be distinguished is high, the biometric samples are poor in quality, an open-set scenario is used, and imposters are introduced to mislead recognition systems. However, in biometric anonymization by contrast, these same conditions do not pose a challenge. In fact, for example the high number of identities makes anonymization much easier, because the more identities we have, the more likely it is that for each identity there is another similar identity in the dataset. This makes it harder to distinguish between identities, which makes anonymization easier. We conclude that anonymization performance will not be accurately evaluated by methods designed to evaluate the performance of recognition systems.

Our analysis shows that the reusing of evaluation methods from recognition and anonymization causes three main problems.

The first problem we identified is that reuse of the scenario for the evaluation of recognition makes for an unrealistically weak adversary model for the evaluation of anonymization. Since in most papers the recognition system is trained on clear data and not on anonymized data (e.g. [177, 140, 301]), obviously the implicit assumption being made is that the adversary is unaware of the anonymization in place. However, an adversary which is aware of the anonymization can adapt to the anonymization and thus will present a greater threat. Consider, for example, an anonymization that performs a deterministic block permutation on a face image. The modification of the data would most likely cause the trained recognition model to break down, and therefore report a high performance. That report, however, will be based on flawed premises and is false.

The second problem we identified is that most evaluation methods assume that the recognition model which works best on clear data will also be the best model for recognizing people in anonymized data (e.g. [351, 23, 89]). We challenge this assumption. Recognition models are developed on clear data. No consideration is given to tampering with the data. Therefore, we doubt whether the recognition model which works best on the clear data is also the best for anonymized data.

The third problem we identified is that the same datasets are used to evaluate anonymization as are used to evaluate recognition (e.g. [177, 223, 301]). Consequently, anonymization techniques are evaluated almost exclusively on large numbers of identities. We argue that it is more challenging for anonymization techniques when there are low numbers of identities in the dataset. Furthermore, a low number of identities is more realistic because biometric data seldom exists alone and additional individuating information (e.g. device ids, soft biometrics, etc.) can be used to further reduce the number of identities in the group.

6.2.2. Our Improvements to State-of-the-Art Evaluation Methods

Now that we have identified the problems, we will explain our improved evaluation methodology and how it addresses the issues.

We use closed-set recognition for our general scenario to have a stronger attacker. Our adversary possesses a list of identities and consequently may simply test samples against the list to select the most likely identity for a given sample. We use two different biometric recognition system architectures for the gait and face recognition systems. For our gait recognition systems, we use an architecture which only uses data specific to the target identities, and for our face recognition systems, we use an architecture that uses additional background data not specific to the target identities (see Fig. 6.1). Both architectures split the samples of each identity contained in the evaluation dataset into *train set* and *test set*. The train set is used to learn a representation for each identity which is then used to infer the identity of the samples in the test set. In addition to this, the face recognition systems are *pre-trained* prior to training on the train set. During pre-training, an additional background dataset representative of the general population is used to learn the features which can be used to differentiate between identities.

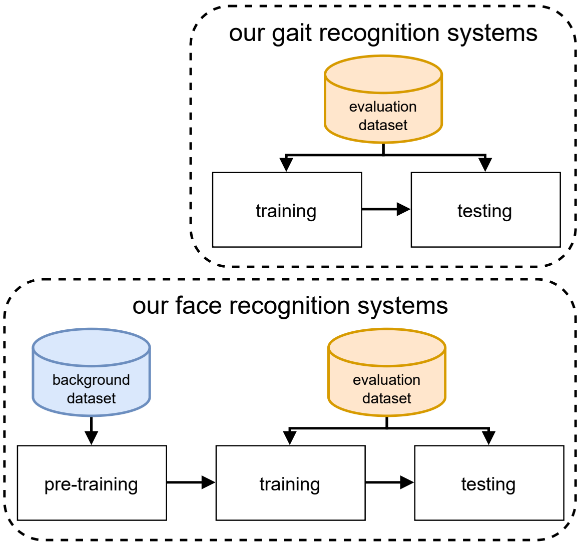


Figure 6.1.: Dataset use through the phases of our recognition systems for face and gait recognition.

Training Recognition Systems with Anonymized Data

In line with previous work [259, 225, 338, 359], we propose that recognition systems be trained on anonymized data so that a more reliable assessment of the anonymization performance is achieved. The idea of retraining recognition systems was first proposed for face recognition by Newton et al. [259]. Their model is trained with anonymized data and then tested on anonymized data. The authors call this scenario parrot recognition, as opposed to training with clear data, which they call naive recognition. The authors report much better performance for parrot recognition compared to naive recognition. Due to training on anonymized data, the biometric recognition system can learn to use features for identification, making this approach more effective than training the system on clear data.

Parrot recognition is another term for an informed attacker, as defined by Srivastava et al. [338]. In the evaluation of voice anonymization, Srivastava et al. [338] propose three attackers who differ in their awareness of the anonymization. The ignorant attacker is unaware of the anonymization (as in black-box assumptions), the semi-informed attacker knows the anonymization algorithm (as in gray-box assumptions), and the informed attacker knows the algorithm plus the given parameters (as in white-box assumptions).

The VoicePrivacy challenge [359, 357] used anonymized data to train a speaker verification system. The system was then tested against anonymized voice samples. It was found that training with anonymized data already improved recognition performance; however, performance improvement was greater when the recognition was pre-trained with anonymized data. The results of the VoicePrivacy challenge show that (pre-)training the recognition system with anonymized data leads to a much stronger evaluation of the privacy performance of a technique. Therefore, we recommend training and (where in use) also pre-training recognition systems with anonymized data. But even when a complete pre-training of the model is not possible, just training with anonymized data can already pose a more difficult challenge to an anonymization.

Test Against Different Recognition Systems

Most evaluation methods rely on the state-of-the-art recognition system currently available for the targeted biometric trait. However, during the design and development of recognition systems, anonymization is not considered. Consequently, recognition systems are not optimized to operate on anonymized data. For this reason, we challenge the assumption that the state-of-the-art recognition systems will also be the one that performs best on the anonymized data. Obviously, for practical reasons, not all types of recognition systems can be used in an evaluation. However, at least a few conceptually different recognition systems should be tested in order to assess which techniques work best on the anonymized data. The aim here is to approximate worst-case performance of the anonymization.

Use a More Challenging Evaluation Dataset

The datasets currently being used for the evaluation of biometric recognition are, as explained, recorded and designed to pose a challenging recognition problem. It is our proposition, though, that evaluators of anonymization use an easy recognition problem in order to create a challenging anonymization scenario. Since the recording of biometric datasets is time-consuming and expensive (not to mention complicated by legal regulations like GDPR), we propose that existing recognition datasets be adapted so that the easy recognition problem becomes a hard anonymization problem. In particular, instead of using the entire dataset, we propose that the identities in the dataset be reduced in number. In smaller groups, it is easier to distinguish individuals because the likelihood of finding similar ones decreases. To further decrease similarity among individuals, we propose basing identity selection on the criterion of easy distinguishability. For the reduced dataset, our identity selection strategies are as follows:

- **Random**: As our baseline selection strategy, we use a random selection of identities. We repeat the selection multiple times to account for the variability of the selection.
- **Classification**: We use a biometric recognition system on the anonymized data to select the identities which have the highest identification accuracy.
- **Metadata**: We operationalize the fact that most biometric datasets also contain metadata about the identities, such as age and sex. Such metadata will typically be extractable via a recognition system. Our rationale is that identities with diverse attributes can be distinguished more easily when images are anonymized. We do this in three steps. First, we normalize each point of metadata information between 0

and 1, and then we calculate the pair-wise Euclidean distance between the points. Second, we obtain a subset of identities by locating pairs of identities at the greatest distances from one another. And third, we calculate the average of distances between the identities in our subset, and then we consistently select the identity located at the maximum distance to the average.

- **Feature-space**: Many recognition systems work by projecting the biometric data into a feature space and then calculating distances between the feature vectors. The rationale is that the recognition system is trained to project datapoints from the same identity onto similar features and as well, to project datapoints from different identities onto contrasting features. However, misclassification occurs when the feature of a datapoint belonging to one identity is farther from the correct feature and closer to a feature belonging to another identity. Therefore, we propose that recognition performance be improved by the intentional selection of identities whose feature vectors are distant from one another on anonymized data. In other words, we choose identities who are very different to one another when anonymized. We use this idea to develop two selection strategies:
  - **Distinctive**: Inspired by the Biometric Menagerie [383], we calculate for each identity a genuine score and an imposter score (illustrated in Fig. 6.2). The genuine score of an identity is the furthest Euclidean distance of any feature vector of this identity to the average of all feature vectors of this identity. The imposter score is the shortest Euclidean distance of the average of all feature vectors of this identity to a feature vector of any other identity. Thus the genuine score is effectively an intra-class distance; conversely, the imposter score is effectively an inter-class distance. If the inter-class distance is high and the intra-class distance low, then the identity is less likely to be misclassified because the features of other identities lie farther away. In sum, we select identities that have the best average of genuine and imposter scores.
  - **Center**: Our purpose is to create a subset of identities lying at the greatest distances from one another. As with the metadata vector above, we begin by selecting the two identities whose average feature vectors have the largest Euclidean distance. Then we consistently select the identities whose average feature vectors lie at maximum distances from the average feature vector of our subset of identities.

6.3. Experiments

Our evaluation is based on the physiological biometric face and behavioral biometric gait. We begin by stating our hypotheses, and then we describe the experiments and present the results.

6.3.1. Hypotheses

Our aim in the evaluation is to test our three methodological proposals for improvements to the evaluation of biometric anonymization. We have proposed, first, that recognition

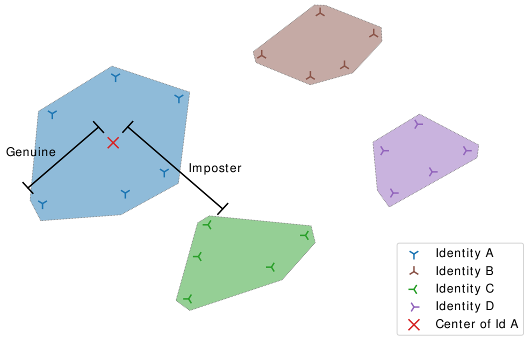


Figure 6.2.: Simplified example for Genuine and Imposter scores of an identity A in a 2D projection of the feature space.

systems also be trained on anonymized data; second, that multiple recognition systems be used; and third, that a more challenging dataset be used.

We begin our testing by formulating five hypotheses:

H1 Training the recognition system on anonymized data achieves more reliable anonymization performance than training on clear data.

H2 Training the recognition system on data in which a part of the samples is anonymized achieves more reliable anonymization performance than training on clear data.

H3 No single recognition system simulates worst-case performance on all anonymizations.

H4 A reduction in the number of identities in the evaluation dataset more robustly challenges the privacy protection of the anonymization.

H5 The identities selected by our selection strategies are a more robust challenge to the privacy protection of anonymization.

Our Hypotheses H1 and H2 hold that training recognition systems on anonymized data will achieve higher recognition performance. For our **H1**, we expect that (pre-)training recognition systems with anonymized data of the respective anonymization will result in higher recognition accuracies compared to (pre-)training on clear data. Further, for **H2**, we expect also that (pre-)training on partial anonymized datasets will perform better compared to (pre-)training on clear data. Further, we expect that increasing the amount of anonymized data in the train set will increase the recognition performance. We reason that the models we test must necessarily generalize more suitably to data that are noisier.

Our Hypothesis **H3** holds that no single recognition system will achieve the best performance on every anonymization. Our prediction for H3 is that, independent of results on clear data, some recognition systems will outperform others when using anonymized data. We reason that some recognition systems will better learn features from the anonymized data.

Our Hypothesis **H4** holds that reducing the number of identities in the evaluation dataset will present a more robust challenge to the performance of the anonymization. Our H5 builds on H4. For **H5**, we expect that selecting an evaluation dataset with our proposed selection strategies will pose a bigger challenge to the anonymization, and hence result in higher recognition performance then selecting random identities.

6.3.2. Experiments

We set an optimal performance bound by using chance-level performance of the anonymization as our baseline. We reason that perfect anonymization would leave adversaries with such a negligible advantage that their most effective strategy would be to guess identities at random. To approximate worst-case performance of the anonymization, we use the performance of clear level recognition, that is, the performance of the recognition system on clear data.

To test H1, we follow the same procedure for each anonymization technique: the recognition system is trained on the respective anonymized training data, and where possible, the system is also pre-trained on the anonymized data. To test H2, we (pre-)train the recognition system on different compositions of anonymized and clear training data using 25%, 50%, and 75% anonymized training data. Hence, we assess our H1 and H2 each with parrot and naive recognition.

For our H3, we use different recognition systems and perform parrot recognition for each anonymization.

For our H4, we again perform parrot recognition. However, instead of using the full evaluation dataset, we use only a random subset of identities of 50%, 25%, 12.5%, ..., until three of the original identities remain. For each number of identities, the sampling is repeated ten times to account for the variability of the random selection. Finally, in our last experiment for H5, we use the same numbers of identities as in the experiments for H4, but instead of randomly selecting, we choose identities according to the strategies described above in our methodology: Random, Classification, Metadata, Distinctive, and Center (see Subsection 6.2.2). We repeat the classification of the reduced dataset ten times to account for the randomness of the test/train split.

6.3.3. Datasets

For the face recognition, we use the CelebA [190] dataset because it is popular for face recognition and for anonymization evaluation, and we use the WebFace260M [410] dataset because its images are realistic. From both datasets we randomly select 1,000 identities as evaluation set and another 9,000 identities as background dataset for retraining. We only select identities with at least eight images, and we limit the maximum number of images per identity to 20. We crop all images to the face region, with images containing multiple faces cropped to the largest face. We resize all images to 224x224 pixel and rotate them until the eyes are level.

For gait we use the dataset of the gait patterns of 57 identities by Horst et al. [134]. The dataset represents the most comprehensive publicly available dataset that contains multiple gait samples per identity, and this, in particular, recommends the dataset to the evaluation of anonymization performance. For each identity in the dataset there are 20 gait sequences, and we resample these to be 100 frames long. The dataset has used optical markers to capture motion. The motion capture covers 52 tracked points, each given as absolute 3D position (see Fig. 6.3).

6.3.4. Evaluation Framework

In order to run our experiments, we implemented the evaluation framework depicted in Fig. 6.4.

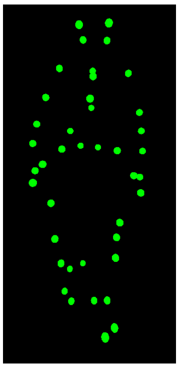


Figure 6.3.: Sample pose of motion-captured gait information, represented as point-light walker.

First, the clear dataset is copied and anonymized with a specific anonymization technique. Second, the selector performs a selection strategy to reduce the dataset to the configured numbers of identities. Third, the splitter splits the samples per identity into two sets, with 75% of samples going into the train set and 25% going into the test set. Depending on the configuration, either the clear samples or the anonymized samples go into the respective datasets.

Fourth and last, the recognition system is trained with the train set and evaluated with the test dataset. The resulting likelihood for a given test sample is recorded and saved for each identity.

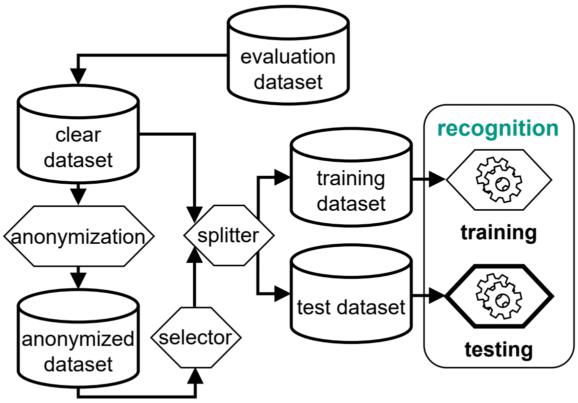


Figure 6.4.: Schematic overview of the evaluation framework architecture, excluding pre-training for simplicity

6.3.5. Recognition Systems

For face recognition, we use the DeepFace [324] library because it covers the entire face recognition pipeline and includes pre-trained models for ArcFace [73], Facenet [323], and VGG-Face [269]. Additionally, we use the face recognition model (frknn) [101], which uses a pre-trained feature extractor and k-nearest neighbors for classification. In order to also test non-deep-learning approaches, we use a scalar, principal component analysis

(PCA) and support vector machines (SVM) pipeline as described in a scikit tutorial[1] and a recognition method that uses Google AI's mediapipe[2] to extract 478 3-dimensional face landmarks before using a scalar, PCA and SVM pipeline on their coordinates. We also pre-train multiple models of ArcFace, which thereafter we referred to as Retrained ArcFace. For ArcFace pre-training, we used the remaining identities in CelebA or WebFace260M with the respective anonymization technique under evaluation applied to the samples. For %-parrot recognition approaches, we anonymized only the corresponding percentage of the samples in the background dataset. We validated Retrained ArcFace on clear data and achieved similar identification accuracy as the regular pre-trained ArcFace.

For gait recognition, we use two types of feature vectors. The flatten feature vector simply flattens all poses of a gait sequence into a single vector, as proposed by Horst et al. [134]. The simple feature vector does a PCA over all poses of a walking sequence and then concatenates the 4 first components of the PCA with an average over all poses of the sequence. For classification, we use SVM, random forest, and k-nearest neighbors. Unless stated otherwise, we used the combination SVM+flatten for gait recognition.

6.3.6. Anonymization Techniques

In the following, we present the anonymization techniques we use for our evaluation. For face anonymization, we select simple anonymization techniques such as blurring and state-of-the-art machine learning anonymizations such as CIAGAN [223]. For gait anonymization, we use a subset of the anonymizations used in Chapter 4. If the anonymization is parameterized, we select the parameters in such a way that initially a low level of recognition accuracy is achieved. In this way, we can observe how our methodological improvements increase the recognition accuracy. Note that since we are investigating the efficiency of our methodological improvements, our selection of parameters does not allow a fair comparison of the anonymizations.

Face Anonymization

For face anonymization we choose a wide variety of different anonymizations, ranging from basic approaches like eye masking to deep learning approaches like DeepPrivacy [140]. For an example of each face anonymization see Figure 6.5. The *Eye Masking* anonymization uses a black strip with 140 pixels height to cover the eye area of the face. *Gaussian Blur* applies a gaussian blur with a kernel size of 101. The anonymization k-randomized transparent overlays (*k-RTIO*) ($\alpha = 0.4$, $blocksize = 18$, $k = 3$) by Rajabi et al.[301] add a block-permuted semi-transparent overlay to the face image. The three methods *DP Pix* [88] ($\epsilon = 2$, $b = 12$, $m = 16$), *DP Snow* [152] ($d = 0.01$), and *DP Samp* [370] ($\epsilon = 5$, $k = 24$, $m = 12$) use differential privacy (DP) to provide formal privacy guarantees. We adapted these three methods from Reilly et al. [305] for RGB images. Our adaptation to RGB images prevents us from providing the formal guarantees given for grayscale images. Another formal privacy framework is k-anonymity, as used in the anonymization *k-Same-Pixel* ($k = 10$) by Newton et al. [259]. k-Same-Pixel expects a static dataset with a single image per identity. This does not apply to our scenario because we anonymize image by image and have multiple images per identity. Therefore, we use a separate background

[1] https://scikit-learn.org/stable/auto_examples/applications/plot_face_recognition.html

[2] https://developers.google.com/mediapipe/solutions/vision/face_landmarker

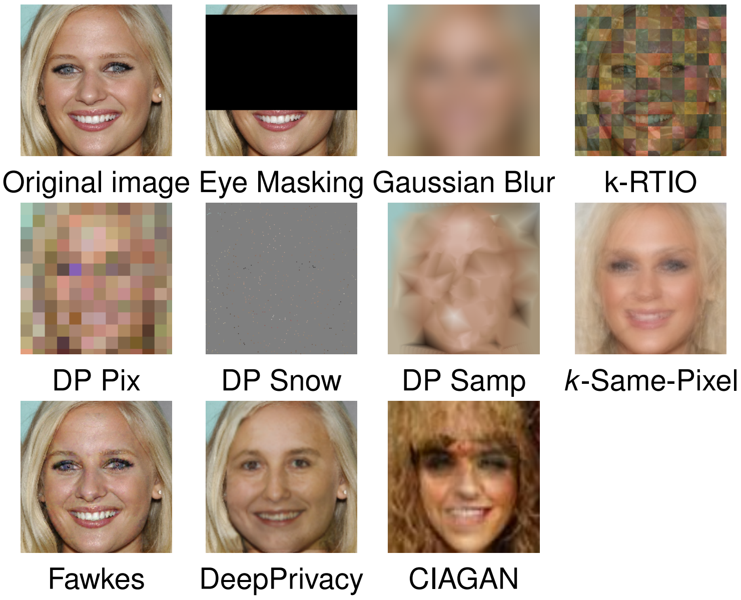


Figure 6.5.: Example image for each of the face anonymization techniques we assess.

dataset with 200 identities. This means that the formal guarantees do not apply to our implementation. In Fawkes [327] $(mode = high)$, adversarial machine learning is used to poison face recognition training data and thereby protect the identity in the picture. Both *DeepPrivacy* [140] and *CIAGAN* [223] anonymize faces by replacing them with new synthetic ones and then fitting them into the original background.

Gait Anonymization

For our gait experiments, we use simple anonymization techniques (see Chapter 4) to select precisely the information to be perturbed in the samples. First, we suppress parts of the samples: *Keep(legs)* and *Keep(head)* both keep only the captured points for legs or head, respectively, while all other points are set to zero. Second, we perturb the samples: *Noise(x)* applies to each captured point normal $(\mu = 0, \sigma = 1)$ distributed noise, which is scaled by 3, 10, or 100. Third, we generalize: *Motion Extraction* captures the differences between each next pose in order to extract only the dynamic parts of the data. The structure of the walkers is, then, effectively removed.

6.3.7. Selection Strategies

For our selections of face data using the Classification strategy, we use ArcFace to calculate the identification accuracy for each identity. We also use ArcFace to extract the feature vectors for the Center and Distinctive strategies. For gait, we use SVM+flatten for the Classification strategy and a PCA with four components over all samples as feature vector for Center and Distinctive.

6.3.8. Framework Implementation

Our evaluation framework was implemented using python (version 3.8) with numpy (1.19.5), scikit-learn (0.23), and DeepFace[324] (0.0.65) libraries.

6.4. Results

We report here the results of our experiments. We assess, in turn, the validity of each of our hypotheses: whether recognition systems trained on anonymized data improve evaluation performance (H1, H2), whether no single recognition system performs best on every anonymization (H3), and lastly whether a reduction in the number of identities (H4) and whether a selection of identities in the evaluation dataset actually pose real challenges to the privacy protection of the anonymization (H5).

6.4.1. Recognition Systems Trained on Anonymized Data Improve Evaluation Performance

In Fig. 6.6 and Fig. 6.7, we present the results of our experiments for H1 and H2 on the anonymization of face data and for gait data.

For face images, we find that, except for CIAGAN and *k*-Same-Pixel, all parrot recognition systems perform better than naive recognition. For *k*-Same-Pixel, all recognition systems have nearly the same performance, while for CIAGAN, naive recognition performs best. This anomaly in CIAGAN makes sense when we consider how CIAGAN performs the anonymization: every face is replaced by another face which shares the same soft biometrics. Therefore, we assume that CIAGAN's replacement of the face on each training image makes it harder for ArcFace Retrained to learn useful feature vectors.

We find significant results for parrot recognition of face anonymization. The performance of full parrot recognition and of all %-parrot recognition cluster close together for most anonymizations. In fact, %-parrot recognition often achieves the same performance as the full parrot recognition, and for DP Snow, the 75%-parrot recognition even outperforms the full parrot recognition.

In contrast to our results for face anonymization, the results for gait anonymization show full parrot recognition outperforming %-parrot recognition, with the exception of all Noise anonymization (cf. Fig. 6.7). For all gait anonymizations, naive recognition performs only at the chance-level. The %-parrot results for Noise(3) and Noise(10) are interesting because 25% performs best, 50% performs second best, 75% performs third best, and full parrot performs worst.

In our results for both face and gait anonymization, one thing defied our predictions. In the face and gait anonymization of DP Snow, Noise(3), and Noise(10) anonymization performance improves when the model is trained solely on a portion of anonymized images rather than on the full anonymized training set. We draw attention to the fact that all three anonymizations perform noise injection either by adding noise to each datapoint or by randomly removing pixels from the image. That portion of noisy data samples in the training set enables the recognition systems to adapt to DP Snow, Noise(3), and Noise(10) while still learning the features required for the classification from the clear data. We conclude, therefore, that there is a tipping point where more noisy data no longer improves training performance but begins impairing it.

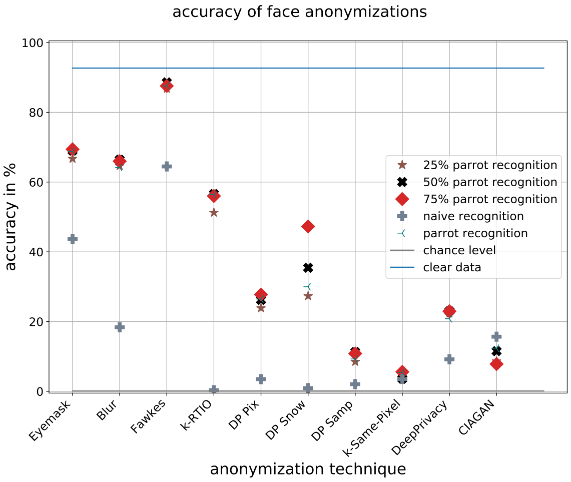


Figure 6.6.: Accuracy for face anonymizations using ArcFace retrained on the CelebA dataset with naive, %-parrot, parrot recognition. A lower accuracy means better privacy protection.

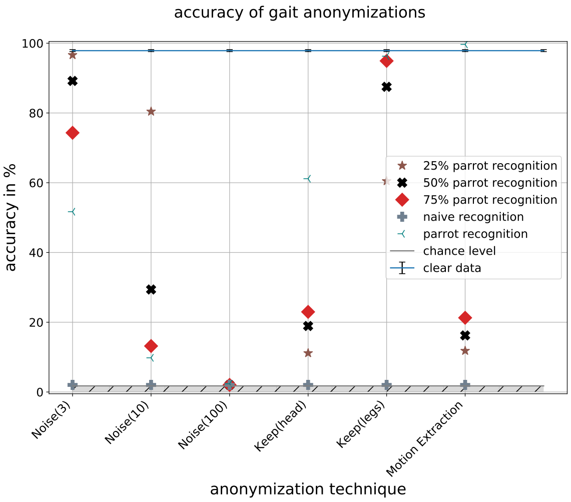


Figure 6.7.: Accuracy for gait anonymizations using SVM+simple with naive, %-parrot, parrot recognition. A lower accuracy means better privacy protection.

6.4.2. No Single Recognition System Performs Best on All Anonymizations

We present the results of our experiments for H3 for the anonymization of face data in Fig. 6.8 and for the anonymization of gait data in Fig. 6.9.

All face anonymizations, except Fawkes, achieve a performance below 30% for all recognition systems except ArcFace Retrained. Fawkes achieves between 30% and 60% (except with Eigenfaces). The results for ArcFace Retrained differ significantly. With ArcFace Retrained, most anonymization techniques achieve much higher recognition rates. Only CIAGAN, DP Samp, and *k*-Same-Pixel are still below 30%, while Blur, DP Snow, and Fawkes are even above 60%. An interesting observation is that while Eigenfaces performs worst on clear data it performs better on DP Pix and Blur than most other recognition systems.

For the gait data, all combinations of techniques perform between 80% and 98% on clear data, with SVM+flatten performing best on the clear data. The gait anonymization techniques across recognition systems perform in the same order, that is, we find the worst performance for Noise(100) and we find the best performance for either Keep(legs) or Motion Extraction.

We note that the differences between the gait anonymization techniques across the recognition systems can be quite large. For example, SVM+simple Noise(100), Noise(10), and Noise(3) score much higher when compared to the other recognition systems. However, among the anonymizations that do not use noise injection, SVM+simple scores lower than SVM+flatten. In sum, we observe that no single gait recognition system outperforms the others.

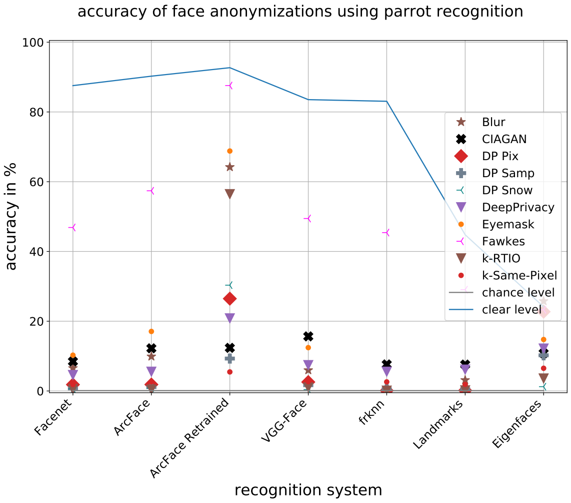


Figure 6.8.: Accuracy of face anonymization over different recognition systems using parrot recognition on the CelebA dataset. A lower accuracy means better privacy protection.

6.4.3. Reducing the Number of Identities in the Evaluation Dataset Increases the Challenge for the Anonymization

We present the results of our experiments for H4 for the anonymization of face data in Fig. 6.10 and for the anonymization of gait data in Fig. 6.11.

For the face data, we assess the accuracy of our H4 by comparing the performances of parrot recognition on different numbers of identities in the evaluation dataset (see Fig. 6.10). For each number of identities (except the number of the full dataset), we selected 10 random subsets and calculated average performance and standard deviation. Every decrease in the number of identities increases the chance-level performance for the recognition systems. In short, the decreases make it easier for the recognition system to randomly guess an identity. We observe this increase in performance for all anonymization techniques. In particular, Fawkes attains the same performance plateau as initially on the clear data. Eyemask, Blur, and k-RTIO also start at high performance, but need longer to approach clear-level performance. *k*-Same-Pixel is the best performing anonymization. *k*-Same-Pixel stays close to the chance-level while mimicking the same increase in accuracy. In

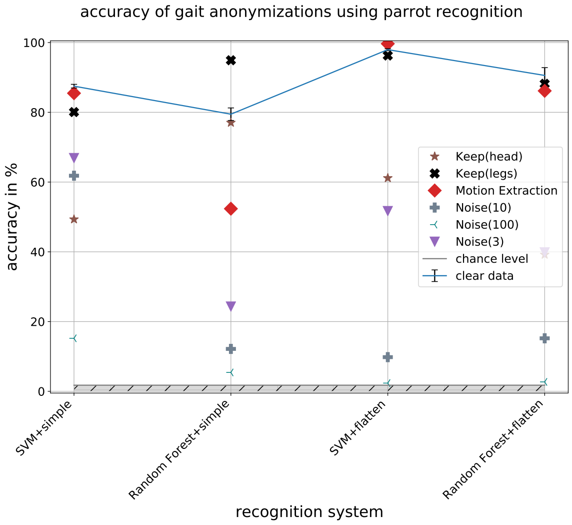


Figure 6.9.: Accuracy of gait anonymization over different recognition systems using parrot recognition. A lower accuracy means better privacy protection.

sum, we observe that decreases in numbers of identities increase the standard deviation of accuracy. From this, we reason that the selection of identities for the evaluation group is an decisive factor in evaluation accuracy.

For the gait data (Fig. 6.11), we observe a similar increase in recognition performance, except for the anonymization techniques Noise(10) and Noise(100), which stay close to the chance-level. The techniques Noise(10) and Noise(100) increase the standard deviation of the performance as the number of identities decreases. For the other gait anonymizations, we do not observe the same relation in the standard deviation.

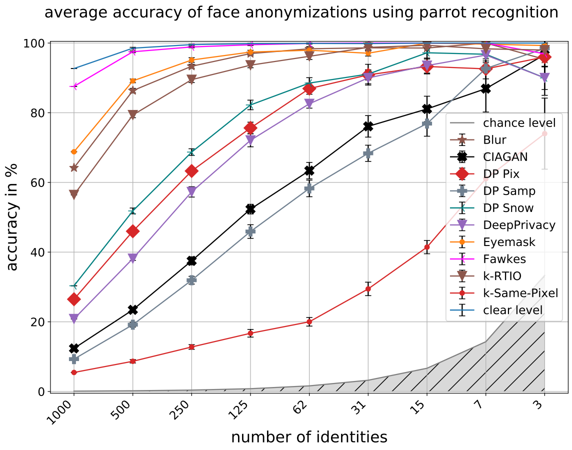


Figure 6.10.: Mean accuracy of face recognition over ten random selections (excluding 1000 identities) from decreasing numbers of identities. The standard deviation of the random selection is given as error bars. ArcFace Retrained is used with parrot recognition on the CelebA dataset. A lower accuracy means better privacy protection.

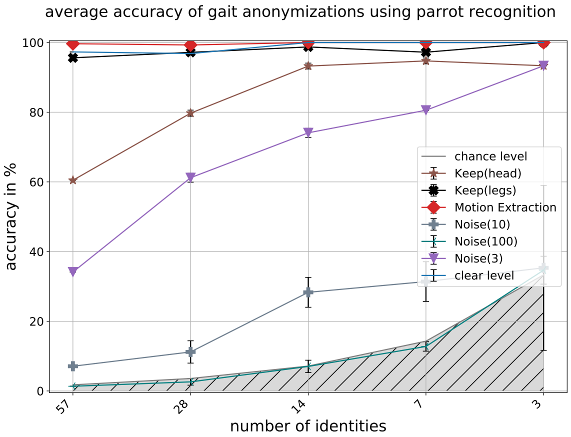


Figure 6.11.: Mean accuracy of gait recognition over ten random selections (excluding 57 identities) for decreasing numbers of identities. The standard deviation of the random selection is given as error bars. SVM+flatten is used with parrot recognition. A lower accuracy means better privacy protection.

We present the results of our experiments for H5 for the anonymization of face data in Fig. 6.12 and for the anonymization of gait data in Fig. 6.13.

Our selection strategies compare to random selection as follows: our strategies outperform when the number of identities is greater than 62, and under 62 Metadata starts performing worse than the best random selections, while the remaining techniques continue outperforming the best random selections down to 3 identities. Our Center and Classification strategies perform best across all numbers of identities, even matching the performance of random selection for 3 identities. What is more, for 500 to 15 identities, our Center and Classification strategies increases over 10% in performance compared to the best random selection.

For the gait data (Fig. 6.13), our results are not as good as for the face data. In general, we find that none of our selection strategies outperforms the best random selections. The strategy that performs consistently best is Classification. It always scores close to the best random selections. The strategy Metadata performs worst, as it does too in the face results. The strategies Center and Distinctive show varying results for different numbers of identities. Our explanation for the contrast between face and gait runs as follows: It is probable that the significant difference between the number of identities in the full face dataset (n = 1,000) and the number in the full gait dataset (n = 57) results in less identities to pick from.

The accuracy we achieve with our Classification selection strategy deserves further attention here, because it performs best across anonymizations for both face and gait. We will examine Classification more closely by comparing it to the initial results for our decreases in numbers of identities.

For the face data (see Fig. 6.14), we observe that clear and Fawkes reach an early plateau close to 100% and that Eyemask, Blur, and k-RTIO begin scoring near the 80% mark and not near the 60% mark. For 125 identities, Eyemask, Blur, and k-RTIO also plateau earlier. DP Samp increases in accuracy steadily from 500 identities to 31 identities, and from there DP Samp accelerates in performance ultimately to achieve 100% at 3 identities. *k*-Same-Pixel achieves the lowest accuracies compared to the other anonymi-

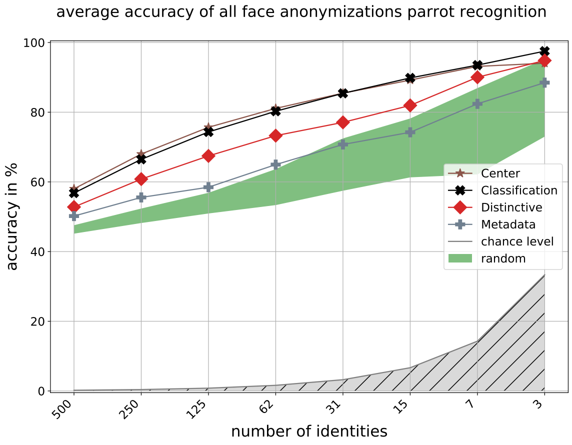


Figure 6.12.: All accuracies are given as the average across all face anonymization techniques. The green area indicates the accuracy range of the previous ten random selections of identities. ArcFace Retrained is used with parrot recognition on the CelebA dataset. A lower accuracy means better privacy protection.

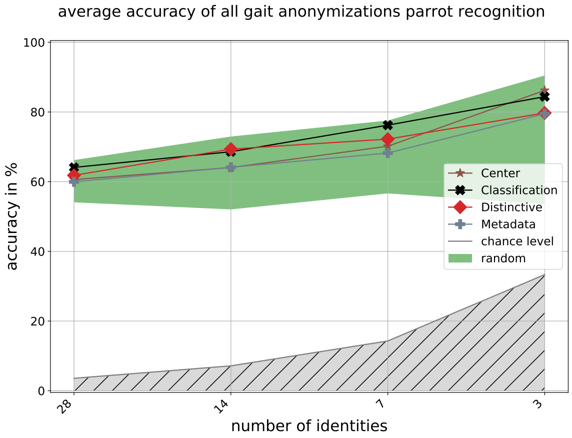


Figure 6.13.: All accuracies are given as the average across all gait anonymization techniques. The green area indicates the accuracy range of the previous ten random selections of identities. SVM+flatten is used with parrot recognition. A lower accuracy means better privacy protection.

zation techniques. However, *k*-Same-Pixel follows the same trend as the other techniques by steadily increasing as the identities decrease in number. When we compare to the random selection (see Fig. 6.10), we see an increase from around 60% to 90% for 3 identities. Similar increases can also be found for the other anonymization techniques. We conclude that the Classification strategy is effective in selecting identities that are hard for the anonymization techniques to anonymize.

For the gait data (see Fig. 6.15), we again find results similar to face. All anonymizations, except Noise(100), score higher. We consider this to be additional evidence that our Classification strategy is highly successful. Furthermore, we find that the Noise(100) results show that anonymization techniques exist which can achieve near perfect anonymization even in this challenging scenario.

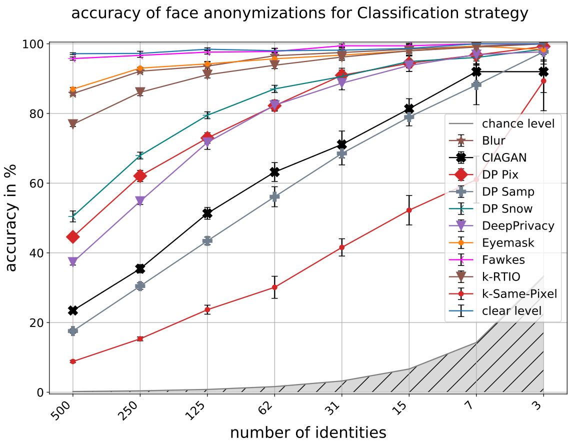


Figure 6.14.: Accuracy of face anonymizations across decreasing numbers of identities. The strategy Center was used to select the identities. The error bars give the standard deviation over 10 test-train-splits. ArcFace Retrained is used with parrot recognition on the CelebA dataset. A lower accuracy means better privacy protection.

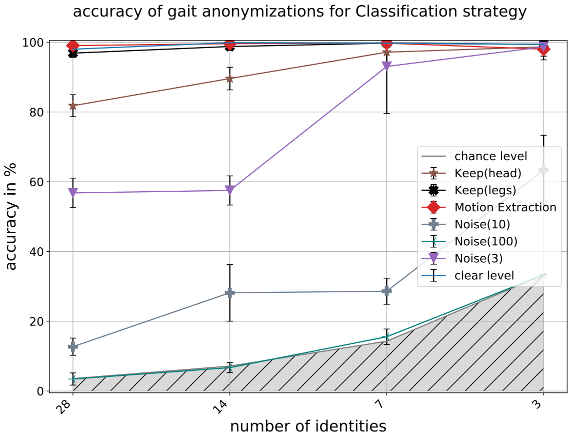


Figure 6.15.: Accuracy of gait anonymizations across decreasing numbers of identities. The strategy Classification was used to select the identities. The error bars give the standard deviation over 10 test-train-splits. SVM+flatten is used with parrot recognition. A lower accuracy means better privacy protection.

6.4.4. Summary of Results

Here, we give a short summary of our most important results:

- Recognition performance increases when the system is trained or especially pre-trained on anonymized data.

- Recognition performance increases when a reduction is made in the number of identities in the evaluation dataset.
- Our Classification selection strategy provides reliable evaluation of anonymization. When, however, the number of identities in the evaluation dataset is very small, Classification might be outperformed by best-case random selections.
- Anonymization techniques perform differently across recognition systems. As a direct consequence, it remains unclear which anonymization technique performs best in conjunction with which recognition system.
- For some anonymization techniques which use noise injection, it is crucial to determine the optimal proportion of anonymized data for both training and pre-training.

6.5. Discussion, Limitations, and Future Work

The results of our three experiments confirm all five of our hypotheses. We see ourselves justified in drawing the overall conclusion that our methodological recommendations will improve the state-of-the-art in the evaluation of the anonymization of biometric data.

Our results for the Hypotheses H1 and H2 clearly show that training and also pre-training with anonymized data significantly improves the performance of the recognition and thus opens the door to improved evaluation of face and gait anonymization. As demonstrated for face anonymization, even a small amount of anonymized data greatly improves the training process. However, our results also indicate that an excess of noisy training data may decrease the performance. Therefore, for anonymization by noise injection (e.g. Laplace mechanism), we conclude that care should be taken to determine the right amount of anonymized training data. Nonetheless, we draw the final conclusion that training with anonymized data significantly improves the validity of the evaluation methodology. Without anonymized data in the train set, the performance of the anonymization is bound to be overestimated.

Our results for the Hypothesis H3 show that the recognition systems which perform comparably to one another on clear data may perform differently from one another on anonymized data. Since the performance on clear data is not a good predictor of performance on anonymized data, we conclude that the recognition system which seems to perform at the state-of-the-art on clear data might not accurately evaluate anonymization performance. This holds especially for anonymizations which use noise injection, as demonstrated by our results for gait anonymization. Therefore, we consider it the minimum that multiple recognition systems be used with different model architectures. Furthermore, we recommend designing recognition systems to be more resistant to anonymization. Our reason is clear: there is no single recognition system that performs best in all cases, not for face anonymization and not for gait anonymization. Understanding which recognition system architecture works best for which anonymization together with training the system on anonymized data will help to achieve more reliable evaluation results.

Our results for the Hypothesis H4 confirm that for most anonymization techniques, a reduced number of identities in the evaluation dataset increases the recognition performance more than what the increase in chance-level can explain. This reduction in the number of identities presents a more challenging scenario for the anonymization. Our results for H4 also show that, as the number of identities decreases, the run-to-run variation of possible results increases. We conclude that the selection of identities for the subset is a significant task in the evaluation of anonymization performance.

Our results for the Hypothesis H5 clearly indicate that a more challenging dataset is generated when our Classification selection strategy is used to select the identities for a reduced evaluation dataset. However, it appears that for very small datasets, multiple random selections can still outperform our Classification selection strategy. Hence, we recommend performing Classification and additionally the random selections in order to determine which performs best at identity selection for the evaluation dataset.

All in all, our proposed improvements will evaluate biometric anonymization techniques much more convincingly than these techniques are currently being evaluated. Further research, however, is clearly necessary. For example, our methodological improvements will need to be validated on other biometric traits. In addition, it remains an open research question precisely which types of recognition systems perform best on which types of anonymization. Answers here will help decide whether, in fact, a systematic approach exists for building recognition systems that perform well on specific anonymizations.

6.6. Chapter Summary

Biometric recognition technologies, such as face recognition systems, pose a real threat to privacy. Therefore, a crucial technique for privacy protection is anonymization, and likewise, evaluation is crucial to anonymization. This chapter assesses the state-of-the-art methodologies used for the evaluation of anonymization techniques, finds flaws in those methodologies, and proposes how the methodologies can be improved.

We find several major flaws in the state-of-the-art methodologies for the evaluation of biometric anonymization. The state-of-the-art evaluation is based on weak and unrealistic assumptions about the adversary. These adversaries act in ignorance of the anonymization in place and are accordingly unable to adapt their recognition systems. These are not realistic adversaries of anonymization techniques. Therefore, the state-of-the-art methodologies largely fail to assess accurately the performance of the recognition.

To begin the work of correcting such flaws, we have proposed to improve the evaluation methodology for the anonymization of biometric data.

- It is our recommendation that recognition systems which are trained not only on clear data but also on anonymized data be used to evaluate anonymization performance.
- Furthermore, we argue that the use of a variety of different recognition systems will improve the rigor of the evaluation. The use of merely a single classifier trained only on clear data might result in unreliable, overoptimistic estimates of anonymization performance. Hence, we recommend using multiple recognition systems trained on anonymized data.
- And lastly, we recommend using a more challenging evaluation dataset to approximate worst-case performance. Our results indicate that such a dataset can be constructed by reducing the number of identities and selecting the easy-to-distinguish identities with our proposed Classification strategy.

We have proposed improvements to the state-of-the-art in evaluation methodologies that will pre-empt overestimations of biometric anonymization performance. We have backed this finding with strong experimental evidence. Thus, we conclude that our proposed improvements lay the cornerstone of a more reliable evaluation methodology for the anonymization of biometric data.

7. Collecting Motion Data For Motion Anonymization

To facilitate our motion anonymization research, we collected **CeTI-Locomotion** [116], a large motion dataset. We captured 50 young people performing different gait tasks using IMU-motion-capture suits. Additionally, during the capturing of the CeTI-Age [287] datasets another 30 people (old and young) have been captured performing the same tasks.

In order to contribute to both the medical and biometric aspects of using motion data, we recorded a new dataset in a young adult sample. Our dataset has the advantages of capturing motion of the entire body with good precision in a relatively large sample of young adults performing four different types of gait movements and multiple repetitions of the sit-to-stand task. Such data allow intra-individual variability and its variations across individuals be taken into account for training machine learning methods. These features are better suited for medically relevant applications for which full body motion capture with a good accuracy are necessary for assessing functional status of different body parts. Further, having a large of number of participants in performing different types of tasks with high numbers of repetitions is important to train machine learning algorithms in identifying biometric features of individuals.

For motion tracking, we utilized an IMU mocap suit as it is a good compromise between setup complexity and tracking quality for full body motion. An IMU sensor consists of an accelerometer, gyroscope, and magnetometer that estimate the position and orientation of the IMU sensor. IMU based motion tracking offers a good accuracy of relative movements of the individual body parts. However, compared to optical marker-based tracking as the gold standard of motion capture, IMU motion capture is less accurate for positions and orientations. The difference between the two approaches is nevertheless slight (e.g., around +/-2 degrees, see Mihcin [239] and later in discussions). An important benefit of IMU based tracking over optical tracking is that it is not limited by a certain spatial volume in which the recording takes place, which makes it easier to capture longer sequences (e.g. Horst et al. [134] capture only a single gait cycle per recording) or to avoid the need of using treadmills (e.g. Troje et al. [360]). In addition, the setup time of a recording session is

This chapter is based on the contributions:

- **Simon Hanisch**, Loreen Pogrzeba, Evelyn Muschter, Shu-Chen Li, and Thorsten Strufe. “A kinematic dataset of locomotion with gait and sit-to-stand movements of young adults”. In: Scientific Data 11.1 2024. DOI: 10.1038/s41597-024-04020-6.
- Loreen Pogrzeba, Evelyn Muschter, **Simon Hanisch**, Veronica Y.P. Wardhani, Thorsten Strufe, Frank H.P. Fitzek, and Shu-Chen Li. "A Full-Body IMU-Based Motion Dataset of Daily Tasks by Older and Younger Adults". Scientific Data 12, 531 (2025). https://doi.org/10.1038/s41597-025-04818-y "The CeTI-Age-Kinematics Dataset: Fully-body kinematic IMU Data of the Elderly and Young Adults in Daily Tasks". 2024. Figshare 10.6084/m9.figshare.26983645.

much shorter, as putting on an IMU suit is much faster than sticking a full set of markers to the bodily landmarks of a participant and the calibration of the IMU system requires also less time.

7.1. Methods

Participants

Data was recorded from 50 young adults (28 male, 22 female; age mean 24.3 years, std. 4.7 years; mass mean 73.5 kg, std. 16.2 kg; height mean 175.5 cm, std. 9.8 cm). An overview of the demographics can be seen in Figure 7.1. All participants gave written informed consent and agreed to the publication of the study data. The study was approved by the Ethics Committee of the Technical University of Dresden (SR-EK-5012021) and was conducted in accordance with the tenets of the Declaration of Helsinki. Data were pseudonymized and a unique ID was assigned to each participant.

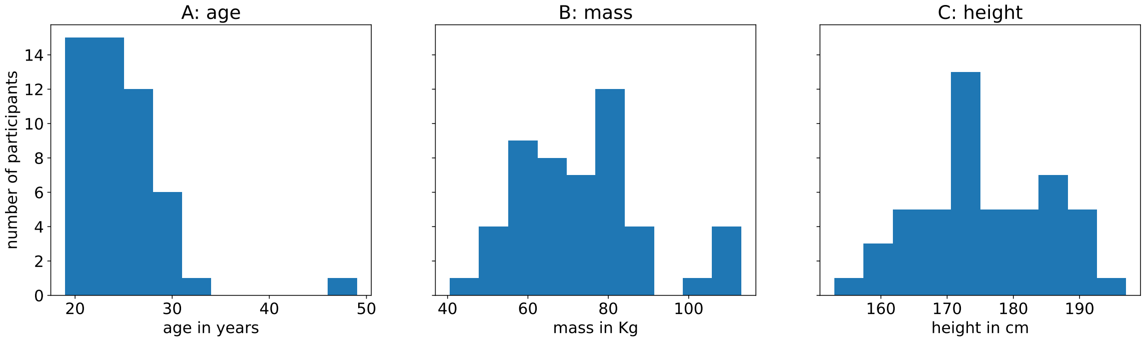


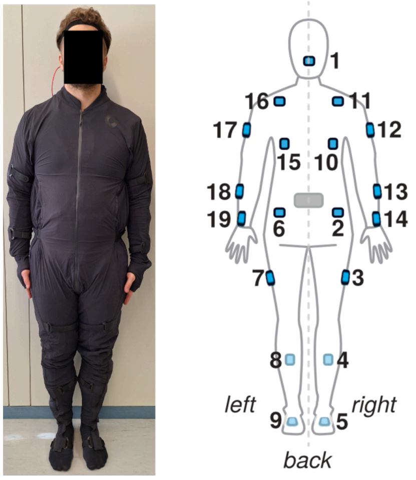
Figure 7.1.: Overview of the demographics of the participants as histograms. Panel A shows the age distribution in years, Panel B shows the mass distribution in kg, and Panel C shows the height distribution in cm.

Acquisition Setup

The dataset was recorded for each participant individually in the laboratory of Karlsruhe Institute of Technology. The single recording session on average lasted for about 60 minutes. Before the recording started, we marked the walking area of 3.5 m with two red crosses on the floor.

Anthropometric Data Anthropometric data was recorded for all participants using an anthropometric grid [121] for total height and arm span of the participants. Additionally, key anthropometric body measurements (i.e., shoulder height, shoulder width, pelvis height, pelvis width, knee height, foot length, and manus length) were taken manually with a generic measuring tape and ruler (stated accuracy ± 0.9 mm) following the anthropometric measurement template of the manufacturer of the *motion capture* (mocap) suits. This data was then used to create a body profile for each participant. Finally, a standard personal scale (Huawei AH100) was used to determine the weight of each participant (wearing a mocap suit and without shoes). The weight values are given with the mocap suit (about 1.2 kg) included.

Kinematic Data Motion data were collected using the Smartsuit Pro 1 (see Figure 7.2 left and middle panel) and Rokoko Studio (version 1.20.5r) mocap technology provided by Rokoko (Rokoko, Denmark, https://www.rokoko.com). The mocap suit contains a total of 19 IMU sensors (see Figure 7.2 right), with one sensor each on the foot, shin, thigh, hand, forearm, upper arm, shoulder, and head. The remaining sensors are located on the torso, with two sensors each on the lower back and hips. The sensors are held in place by Velcro and the fabric of the suit. Mocap data is recorded at 100 Hz and transmitted via WiFi to the recording computer and then applied to the participants' skeletal rig. The skeletal rig is a representation of the key body parts with the proportions of the participants' body profile (described above).



Number	Sensor Location	Segment Label	Joint Label
1	head	Head	Neck
2	right pelvis	Pelvis	Pelvis
3	right upper leg	RightUpperLeg	RightHip
4	right lower leg	RightLowerLeg	RightKnee
5	right foot	RightFoot	RightAnkle
6	left pelvis	Pelvis	Pelvis
7	left upper leg	LeftUpperLeg	LeftHip
8	left lower leg	LeftLowerLeg	LeftKnee
9	left foot	LeftFoot	LeftAnkle
10	right chest	Chest	Thorax
11	right shoulder	RightShoulder	RightScapula
12	right upper arm	RightUpperArm	RightShoulder
13	right fore arm	RigthForeArm	RightElbow
14	right hand	RightHand	RightWrist
15	left chest	Chest	Thorax
16	left shoulder	LeftShoulder	LeftScapula
17	left upper arm	LeftUpperArm	LeftShoulder
18	left fore arm	LeftForeArm	LeftElbow
19	left hand	LeftHand	LeftWrist

Figure 7.2.: Participant wearing Rokoko Smartsuit Pro 1 in calibration pose (left) and schematic sensor locations (middle and right) on the suit depicted from a posterior view. The sensors are color-coded, with light blue sensors (4, 5, 8, 9) positioned anteriorly. The segment and joint labels, provided by Rokoko Electronics Inc., are named according to the tracked centers of body segments, with rotation axes conforming to the standardized Joint Coordinate System (JCS) defined by the International Society of Biomechanics (ISB) [376, 377]. A full overview of the channels of each recording can be found in the *_channels.tsv files.

Acquisition Procedure

Figure 7.3 offers a schematic overview of the study protocol. Before the data acquisition started, participants were asked to change in a thin layer of sports garments, over which they wore the mocap suit. A private changing room was provided. Participants were asked to perform the exercises without shoes, wearing socks only. The mocap suit's sensor locations on the limbs were carefully adjusted to the participants' individual body morphologies according to the manufacturer's guidelines. Next, participants were familiarized with the mocap suit and technology and given a general overview of the types of movements they had to perform. They were informed about how the movements and anthropometry of their body would be recorded. We then proceeded with the collection of anthropometric measurements.

The data acquisition started with the actor profile set up in Rokoko Studio where participant-specific ID, demographics, and anthropometric measurements were recorded and the system was calibrated. The calibration pose (straight pose with legs, arms, hands and fingers straightened, feet hip-wide apart) was used to set up the initial sensor position to ensure correct motion tracking. The neutral pose (stand upright with upper body straight, feet foot-width apart, weight evenly distributed on both feet, arms hanging loosely) served as a neutral reference pose from which the movements for all walking exercises were initiated. While the calibration pose places great emphasis on the correct alignment and rotation of body segments, the neutral pose should reflect a natural and comfortable resting pose. Participants then performed a series of control movements covering the range of motion of multiple body parts (e.g., shoulder abduction and flexion, wrist flexion and extension, ankle plantar flexion and dorsiflexion) to ensure best mocap quality. If necessary, sensors were re-positioned and the calibration was restarted. For two participants (sub-K15 and sub-K35) minor recording errors persisted for their arms, as specified in the metadata of the participants.

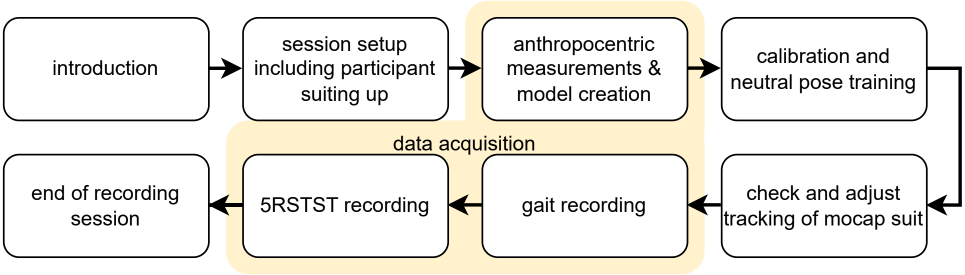

introduction
session setup including participant suiting up
anthropocentric measurements & model creation
calibration and neutral pose training
check and adjust tracking of mocap suit
gait recording
5RSTST recording
end of recording session
data acquisition

Figure 7.3.: The study protocol.

Next, the participants were guided through the recording session, starting with the sequence of gait trials and finishing with the 5RSTST trials, in a predefined order (see also Table 7.1). Prior to each task, a mandatory calibration procedure was conducted to ensure accurate motion tracking. In order to capture potential intra-individual and inter-individual differences in gait movements, our comprehensive dataset includes kinematic data collected under four distinct conditions of walking. The first condition involved recording the natural walking speed and gait style of each participant at their preferred pace. In the second condition, participants were instructed to imagine being in a hurry, simulating a fast walking scenario. The third and fourth conditions focused on walking with an additional load at the participant's preferred speed. Participants either carried a 5 kg backpack (total weight) or transported a standard bottle crate (measuring 400 mm x 300 mm x 270 mm, with a total weight of 5 kg evenly distributed). For all gait movements, participants were provided with explicit instructions to initiate from a neutral pose and walk between two points marked on the ground with red tape, positioned 3.5 m apart (see Figure 7.4 left). Additionally, participants were instructed to maintain a forward gaze aligned with their walking direction. Participants were given specific instructions to execute a controlled turn at the marked points before commencing the return walk. Each gait task consisted of five back-and-forth walks between the designated points. The experimenter verbally counted the number of trials completed during each task. In the final round, participants were instructed to execute an additional turn, ensuring they returned to the original starting position and faced the initial direction.

Following the completion of the walking sequence, data collection proceeded with the *5 times sit-to-stand test* (5RSTST) trials (see also Figure 7.3). The sit-to-stand test is a widely used motion test that should reveal significant differences between age groups. Participants were provided with specific instructions to sit on a chair, crossing their arms in front of their chest, and perform rapid repetitions of standing up and sitting down (see Figure 7.4 right) five times consecutively, without any intervals of rest. The execution time of each 5RSTST exercise was measured using a stopwatch, while the kinematic data was simultaneously captured using a mocap suit. Each participant performed this exercise twice.

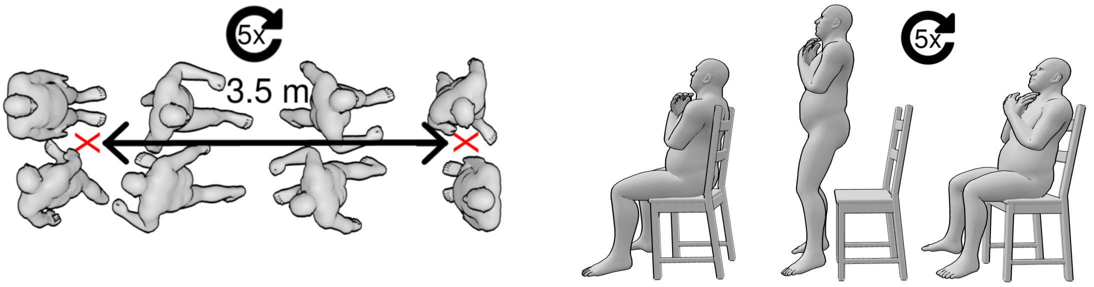


Figure 7.4.: Schematic representation of a gait trial (left) shown from the top and 1RSTS trial (right) shown from the side.

Data Processing

As part of the Rokoko Studio data recording pipeline, the position and orientation data was processed using the following filters. This involved the locomotion filter for all recordings that automatically aligns the feet sensor to the ground plane for movement phases with (assumed) ground contact. Additionally, a drift filter was applied for all gait trials to correct for position drift in the gait trajectories. Since IMU sensors only record the relative motion of each body part, the global position of the suit is estimated by summing up all relative movements beginning from the initial position (established during calibration). Over recording time, this method can lead to increasing positional inaccuracy, known as global drift. Since the start and end point of the walking course in the gait tasks was at the same position, this information was used to correct for global drift. Data was then exported in CSV format. The data includes the absolute positions (x, y, z) of the center of mass of all 17 body segments in meters and the relative orientation of each joint as recommended by the International Society of Biomechanics (ISB) [376, 377]. The ISB standard defines the local axis system of joints in degrees with two fixed axes and one "floating" axis. In total, the data contains 17 6-dimensional (6D) points per time frame.

Finally, the data was processed using a custom Python script [116] (see Figure 7.5). The data was visually inspected for noisy data segments and sensors. Anomalous sensor behavior was identified at individual time points within the position data of two participants (sub-K8, sub-K58), characterized by abrupt signal increases in multiple orders of magnitude. To remedy these artifacts, an interpolation approach utilizing the neighboring data points was implemented. Subsequently, we proceeded with cutting the recordings into individual segments. For the gait task a segment is a single gait cycle (as defined by Perry et al. [145]), and for the 5RSTST task it is a single 1RSTS repetition. A 1RSTS repetition begins with a participant sitting upright in a chair and ends with the participant reaching the same sitting position after standing up. This segmentation step allowed for a more granular analysis and examination of specific movements within each segment.

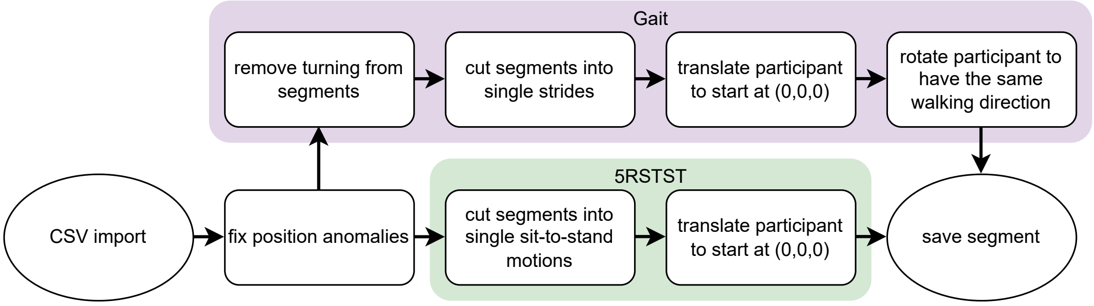

Gait
remove turning from segments
cut segments into single strides
translate participant to start at (0,0,0)
rotate participant to have the same walking direction
5RSTST
CSV import
fix position anomalies
cut segments into single sit-to-stand motions
translate participant to start at (0,0,0)
save segment

Figure 7.5.: Data processing pipeline in custom Python script.

Data segmentation was based on the characteristics of the positional data signal. For tailoring the segmentation approaches to the characteristics of the different tasks, selected positional values of the signal were processed and searched for distinct events, e.g., peak values, to subsequently crop the signal into segments. The positional values are mapped to a coordinate system with the starting point of the movement is set to the coordinate origin (set during calibration), with the x-axis reflecting the position in the transverse plane (e.g., shoulder positions left and right to the median plane), the y-axis describing the heights of body parts (e.g. the head height), and the z-axis displaying the translation in the median plane (e.g., a forward step). While straight walking, the distance between the shoulder x positions (reflecting the approximate width of the shoulder girdle) stays constant. The distance changes at the turning points of the walking course as the participant rotates around its own body axis, from positive to negative or vice versa.

The resulting segments were then cut into single gait cycles, i.e., the feet end in the same position as they started. We chose gait cycles as they contain all gait phases performed while walking. The gait cycles were identified by detecting the largest distance between the absolute z-positions of the participants' feet. As a final processing step, we normalized all gait cycle segments to start at the same coordinate point by setting the positional data of the hip at the starting frame of the segment as the new coordinate origin of the coordinate system (with translating all positional values accordingly). Next, we applied a rotation step to the positional values of each segment so that all segments reflected the same walking direction. This processing step was necessary, since the participants were walking back and forth at the walking course, resulting in different walking directions and corresponding positional values.

For the 5RSTST, we identified a single segment by observing the participant's head position. We detected peak values in the (y, z) values of the head position, since these can be used to distinguish the sitting from the standing position. We automatically sliced them into five repetitions (integrating the standing and sitting phases into one repetition). As with the gait data, we used the first hip position of each segment as the coordinate origin for this segment.

An overview of the resulting number of segments processed can be found in the Table 7.1. Note that the resulting amount of segments for each participant and task varies due to several factors: faster walking results in fewer segments per participant, individual physiology affects the number of segments per person, and because we excluded incomplete gait cycles we also remove incomplete ones which increases the differences between participants. For 5RSTST, we also found an incorrect execution of the task, as one participant performed more repetitions than instructed.

Table 7.1.: Overview of the resulting number of segments (N) per task after processing.

task	N_{total}	N_{average} per participant	N_{min} per participant	N_{max} per participant
gait normal	1178	23.56	12	35
gait fast	846	16.92	9	23
gait bottle crate	1120	22.4	15	33
gait backpack	1029	20.58	11	30
1RSTS	499	9.98	9	11

7.2. Data Records

We provide the CeTI-*Locomotion* dataset [116] on the figshare data exchange platform (https://figshare.com/). The dataset is formatted according to the BIDS [104, 149] standard (version 1.9) and is provided in both raw format (as exported from Rokoko Studio) and processed format (as described above). The repository contains a readme, license, BIDS dataset descriptor, and the Python code used for processing and technical validation. The *participants.tsv* file stores the metadata for all participants, including their age, mass, height, anthropometric data, and 5RSTST completion time. Each subject folder contains the raw motion data for all tasks. The motion data is stored as a tab-separated values file (**_motion.tsv*) accompanied by a metadata file (**_motion.json*) and a channel descriptor of the motion file (**_channels.tsv*). The processed data is located in the derivatives folder and follows a similar structure, but split into the processed segments. See Appenix B.1.2 for the coda availability and Appendix B.1.1 for the usage notes.

7.3. Technical Validation

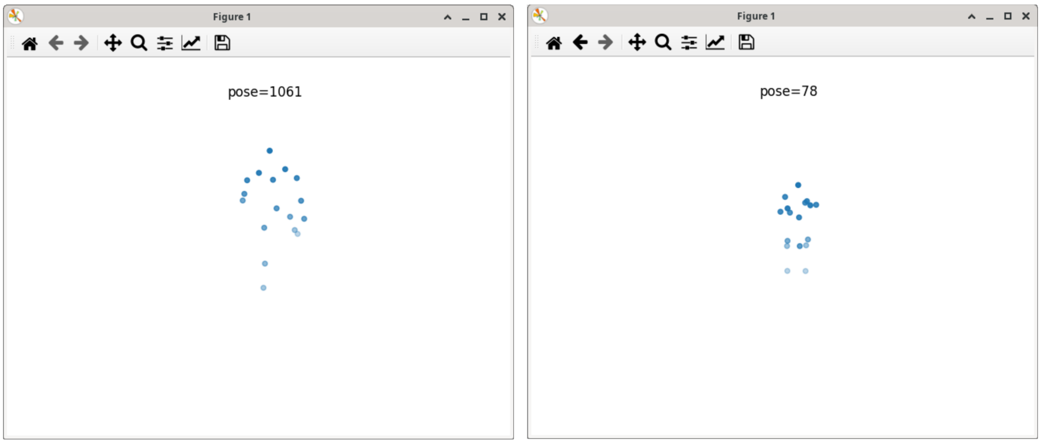


Figure 7.6.: Example of a visual inspection rendering of a gait (left) and 5RSTS (right) recording.

In general, the suitability of using the Rokoko Smartsuit for motion data recording has been validated by previous work. Mihcin et al. [239] reported good agreement between the Smartsuit and an optical tracking system, with a maximal bias range -1.48 – 2.22 degrees for knee flexion and for hip abduction-adduction. We performed two types of validation, one prior to recording our data. Here, we visually inspected the tracking performance of the Rokoko Smartsuit by having participants perform static control poses, such as a T-pose (with 90 degrees shoulder abduction) or lifting their legs. Only when we found the tracking to be accurate, we did start recording the mocap data. The second validation was done after exporting and processing the data into Python. In a first step, we rendered the position data of a random selection of segments to visually inspect a natural-looking motion execution within the anatomically possible range of motion (ROM) (see Figure 7.6). Following this initial manual inspection, we evaluated whether the minimal and maximal joint angles fell within the anatomically feasible ROM. For each participant, we first calculated the minimal/maximal joint angles across all motion segments, for each of the six raw recordings separately (5RSTS is split into two recordings). We then averaged the minimum and maximum joint angles for each participant across recordings. Figure 7.7 depicts the distribution of mean minimal (in blue) and mean maximal (in orange) joint angles for all participants in comparison to reference range of motion values from Ryf and Weymann [316] (in gray). Note, that we only plot joint angles for which reference values are available. For most joints the values fall within the expected ROM. Outliers in knee flexion, hip external/internal rotation, and ankle inversion/eversion could be attributed to inaccurately executed calibration poses or faulty sensor measurements (e.g., for elbow flexion in sub-K9). Additionally, joint angles for wrist abduction exceeded the normal ROM for some participants, likely due to inconsistent positioning of the wrist during calibration.

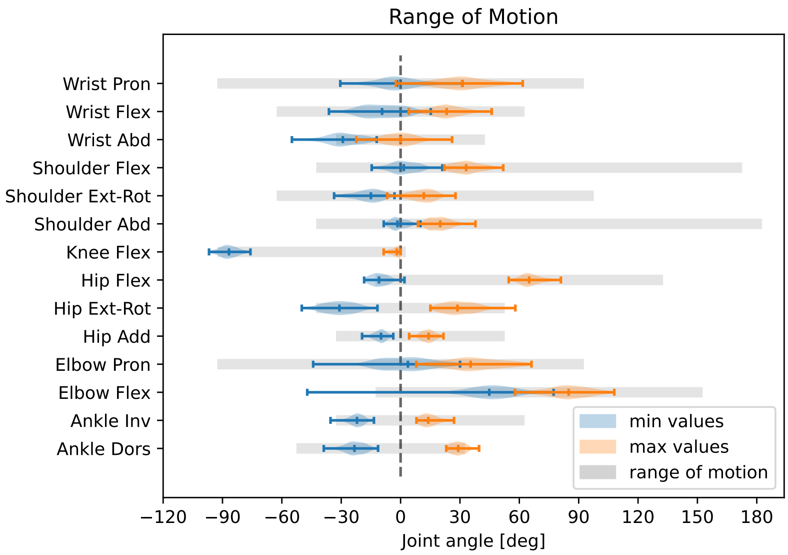


Figure 7.7.: Distribution of average minimum and maximum joint angles for all participants, as calculated across all raw recordings per participant. Anatomically typical range of motion (ROM) values reported in prior research [316] are shaded in gray.

In addition, a classification analysis was performed to assess the suitability of the processed data for machine learning applications. Our model setup is comparable to Horst et al. [134] and we have provided the our source code as part of this Data Descriptor (see Section Code Availability). For classification, we employed a support vector machine (SVM) [60, 322]. The SVM was utilized to classify the movement tasks, the sex of the participants, and the participants' identities. We chose a SVM with a radial basis function (RBF) kernel, a regularization parameter (C) of 1, gamma which is $1/\,(\text{n\_features} * Var(X))$. As a preprocessing step for machine learning, we standardized the length of each motion segment by resampling them to contain exactly 100 frames. Additionally, we concatenated the time frames, which consisted of 17 6D points, into a single vector representation, resulting in a vector of size 10200. To further prepare the data for machine learning algorithms, we applied scaling techniques to normalize the values within each dimension of the data. By scaling the values to a range between 0 and 1, we ensured that all features contributed equally to the learning process, regardless of their original magnitude. To reduce the dimensionality of the dataset and capture the most relevant information, we employed principal component analysis (PCA) [277]. This technique transformed the high-dimensional feature space into a lower-dimensional representation while preserving the most significant variation in the data. This resulted in a segment vector of 3342 to 3363 elements (depending on how the split was performed), with the first 10 PCA components explaining 69%-68% of the variance. The kinematic data was split into a test (20%) and a training (80%) dataset. For the identification classification we split segment-wise, i.e. 80% of a participant's segments are in the training dataset and the remaining 20% are in the test dataset. For sex and modality classification, we split participant-wise, with all segments of a participant in either the training or test dataset, with 80% of the participants in the training and 20% in the test dataset. On the training dataset we performed a stratified 10-fold cross-validation using balanced accuracy (mean of the recall per class) as the metric to select the most appropriate model before testing it on the test dataset.

The metrics accuracy (number of correct predictions divided by total predictions) and F1-scores [91] were used to evaluate the performance of classification models and are reported in Table 7.2. As both metrics achieve comparable results we only discuss the accuracy in the following. Our classification model achieves 97% identification accuracy, a 81% sex recognition accuracy, and a 84% action recognition (gait normal: 90%, gait fast: 93%, gait backpack: 88%, gait bottle create: 97%, and 1RSTS: 100%). For action recognition, a similar result can be found when we plot the actions in two dimensions using t-SNE [366](perplexity 30), see Figure 7.8. t-SNE is a stochastic method for visualizing high dimensional data as two dimensional data points while preserving the similarity of the data points. The clusters of the 1RSTS and gait bottle crate segments are clearly distinguishable from the remaining task segments. The clusters for the remaining task segments overlap and cannot be clearly separated, showing their similarity. Note that for each gait modality we see two clusters. The plot is in line with the results of the classification experiments, as 1RSTS and bottle crate gait are easily separated from the other modalities due to their uniqueness, while the separation of the remaining gait modalities is more difficult.

7.4. Related Datasets

Comparing our dataset (cmp. Table 7.3) with other IMU-based datasets, we have a high number of tracked body points. Furthermore, our number of participants, tasks, and samples are in the middle between the high and low number datasets. The category where we score the lowest compared to the other datasets is the total recording time, as we only

Table 7.2.: The classification results given as accuracy and F1-score for different attributes, we also report the number of classes (C) and number of segments per class (N) used for the classification.

classification	accuracy	f1 score	C	$N_{average}$	N_{min}	N_{max}
identity	97%	97%	50	93.44	61	129
sex	81%	81%	2	2336	2046	2626
action all	84%	85%	5	934.4	499	1178
1RSTS	100%	100%	2	2336	499	4173
gait normal	90%	83%	2	2336	1178	3494
gait fast	93%	78%	2	2336	846	3826
gait backpack	88%	72%	2	2336	1029	3643
gait bottle crate	97%	94%	2	2336	1120	3552

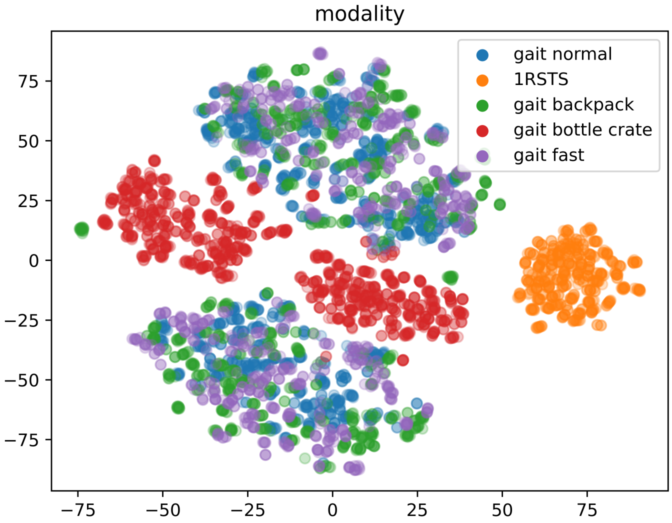


Figure 7.8.: t-SNE plot colored for different motion tasks.

Table 7.3.: Overview of related gait datasets using IMU sensors for motion tracking, sorted by publication year
.

Source	IMU Tracked Points	Participants	Tasks	Samples	Recorded Time	Publication
Khandelwal et al. [162]	4	20	7	51	6h	2017
Chereshneve et al. [54]	6	18	12	2111962	10h	2018
Truong et al. [362]	2	230	1	40.000	8.5h	2019
Loose et al. [195]	3	108	6	1080	30h	2019
Luo et al. [202]	6	30	9	1710	8h	2020
Losing et al. [199]	17	20	3	180 - 300	9h	2022
CeTI-Locomotion [116]	**17**	**50**	**5**	**4672**	**1.5h**	**2024**

record 1.5 hours in total. The comparison shows the trade-off we chose for these datasets, instead of maximizing a single category of the dataset, we decided to spread our resources across all categories to get a good variety in all of them.

7.5. Chapter Conclusion

In this chapter, we presented the CeTI-Locomotion datasets. Due to its large number of repetitions and high-quality data, it is an ideal dataset for studying the identification of individuals from motion data. We demonstrated that individuals can be identified with 97% accuracy. This further illustrates that motion data is a privacy-sensitive behavioral biometric factor that must be protected. The collected data will be a valuable resource for future research on identification risks and privacy protections for human motion data.

8. Pantomime: Motion Data Anonymization using Foundation Motion Models

In this chapter, we propose **Pantomime**, the first anonymization for full-body motion sequences that is robust against re-identification while enabling high utility. Pantomime uses foundation motion models to hide the identity of people captured in motion data, which makes it applicable to different motion data formats and eliminates the requirement to train on the data it should anonymize. The use of foundation motion models allows Pantomime to project the motion data into the motion space. Pantomime then anonymizes the motion data in the motion space by adding random noise to it before decoding the motion data back into its original space.

The advantage of this approach is that by adding noise to the motion sequence in the motion space, we create a new plausible motion that is similar to the original. By increasing the noise, the new motion can be further away from the original motion, and thus anonymization can be increased at the expense of motion utility, allowing us to configure the privacy-utility tradeoff of Pantomime.

Furthermore, using foundation motion models has the advantage that they are trained on general motion data and therefore generalize well to different motions [306]. Because of this, Pantomime does not require the enrollment of specific users or the tuning of the model to specific datasets, as in previous work [251]. Furthermore, Pantomime is applied to the motion data time-step by time-step, making it flexible in its application to motion data (e.g., when animating an avatar in a MR chat application).

We investigate the privacy-utility tradeoff of Pantomime by measuring the utility via naturalness and by comparing the similarity of the anonymized sequence to the original using a user study. To measure the privacy protection we measure the person identification accuracy with a state-of-the-art biometric recognition system [134].

Furthermore, we investigate how much of the individual components (e.g., body shape, joint rotations, etc.) of the motion data contribute to person identification. To better understand which component requires anonymization. The contributions of this chapter are as follows:

- We propose Pantomime, the first general technique for anonymizing motion data that does not require training on the dataset to be anonymized and can configure its privacy-utility tradeoff.
- We evaluate Pantomime's privacy protection using two full-body motion capture datasets.

This chapter is based on the contribution:

- **Simon Hanisch**, Julian Todt, and Thorsten Strufe. "Pantomime: Towards the Anonymization of Motion Data using Foundation Motion Models". 2025. DOI: 10.48550/arXiv.2501.07149

- We conduct a user study to investigate the naturalness and action similarity of the anonymized motion sequences.

8.1. Terminology & Background

In the following, we introduce the SMPL body model, and foundation models used in this chapter. A **foundation model** is, as defined by Bomasan et al. [37]: ".. any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks". In this chapter, we adapt motion models trained on a broad motion corpus for the encoding and decoding of motion sequences for the task of anonymization.

8.1.1. SMPL Body Model

The *Skinned Multi-Person Linear Model* (SMPL) [196] model is a body shape model that uses blend shapes to represent the human body shape for different poses. The SMPL decomposes the body shape into a fixed identity-based body shape and a variable shape that depends on the body pose, which is represented as joint rotations. Due to this decomposition into static and dynamic body pose, the SMPL model respects the body shape deformation that occurs in different poses, i.e. soft tissue deformations when a person is moving compared to when the person is standing still. The SMPL model has been trained to minimize the reconstruction error of high-resolution 3D body scans.

For the remainder of the chapter, the SMPL model will be defined as the differentiable function $M(r, \Phi, \Theta, \beta)$ which maps the root translation $r \in \mathbb{R}^3$, root rotation $\Phi \in \mathbb{R}^3$, body pose as joint angles $\Theta \in \mathbb{R}^{3\times21}$ and identity-based body shape parameters $\beta \in \mathbb{R}^{16}$ to the vertices $V \in \mathbb{R}^{3x6890}$. The joint positions $J \in \mathbb{R}^{3\times22}$ can be calculated from the vertices by using a regressor matrix.

To find the SMPL model representation of a given body pose, we perform *3D body fitting*. The goal of the fitting is to find the best parameters for the SMPL model to fit the given data of the body pose (e.g. all limb positions of a person).

8.1.2. Foundation Models

In the following, we will introduce the two basic motion models VPoser and HuMoR that we use for Pantomime.

VPoser

VPoser [276] is a VAE that has learned the probability distribution of plausible human body poses. It takes a single body pose (represented as the 21 SMPL joint angles Θ) as input and tries to output the identical body pose. Since it models plausible poses, it can be used to judge whether a given pose is plausible or not. This is used to fit the SMPL body model to given motion data by using the model as a loss.

VPoser was regularized with a Kullback-Leibler divergence during training using a normal distribution as prior. Hence, the individual dimensions of VPosers latent space a normal distributed ($\mathcal{N}(\mu, \sigma)$). The VPoser encoder takes body poses as joint angles Θ as defined by the SMPL body model and outputs a latent code z.The latent code is then mapped back into the original pose space by the decoder to get the reconstructed pose $\hat{\Theta}_t$.

HuMoR

HuMoR [306] is a CVAE that has learned the transition from one pose to the next. Unlike VPoser, it uses two successive poses as input to its encoder and then tries to output the transition from the first to the second pose with its decoder. With its focus on pose transitions, HuMoR effectively models the distribution of human motion. Like VPoser, HuMoR can be used to fit body models to motion sequences.

HuMoR represents the state of a moving person as a matrix $x \in \mathbb{R}^{69}$ consisting of root translation, root orientation, body joint angles, body joint positions, and the velocities of root translation, root orientation, and joint positions. The encoder of HuMoR takes two sequential states x_t and x_{t-1} as input and generates a latent code z_t. The decoder of HuMoR takes the latent code z_t and first pose x_{t-1} as input and outputs the reconstructed transition Δ from x_{t-1} to x_t. Using the reconstructed transition Δ we get the reconstructed pose $\hat{x}_t = x_{t-1} + \Delta$.

8.2. Related Work

Below is an overview of research in the emerging field of motion anonymization that attempts to prevent identification. We categorize the works according to the type of motion data they anonymize.

8.2.1. Body Shape anonymization

Sattar et al. [319] investigated the privacy of body shapes extracted from single images. They show that a person's body shape is considered private information by performing a small user study, and propose an adversarial perturbation to prevent the automatic extraction of shape information from images. While this work is not motion anonymization, it highlights the need for anonymizing the body shape information that is implicitly contained in motion data.

8.2.2. Motion Capture Anonymization

Malek-Pdjaski and Deligianni [212] developed an anonymization technique for 3D motion capture that extracts features that do not allow identification but can still be used for affect recognition. They attempt to separate the information needed for affect recognition from the information used for identification by using two AE. One AE is trained to be subject-specific and one AE is trained to be affect-specific. The disadvantage of this approach is that the AEs have to be trained on the dataset to be anonymized. Moon et al. [244] proposed an adversarial anonymization scheme for 3D motion capture data in which a machine learning model is trained to minimize the identity recognition and maximize the action recognition. Both approaches are not suitable for preserving the naturalness of the motion data (as determined by a user study), since the benefit must be quantifiable so that it can be used as a loss in the training of these approaches.

Simple Anonymizations Moore et al. [245] suggest using only the velocities of the motion sequences and Miller et al. [240] suggest using only the joint rotations to reduce identifiability. Meng et. al. [231] additional propose adding noise to the joint rotations.

Mixed Reality Anonymizations Nair et al. investigate how the (motion) data collected by VR headsets can be anonymized. They proposed a framework called MetaGuard [253] which claims to protect various attributes collected by VR headsets. Their second proposal *Deep Motion Masking* (DMM) [251] is a machine learning approach which reduces identity similarity while maintaining action similarity. The evaluation of the approaches performed in [251] shows that DMM is effective and can anonymize motion sequences of different datasets against different attackers, but MetaGuard is not effective and does not provide sufficient protection. The drawback of DMM is that it requires a large training dataset and is application specific for the data it was trained on.

8.2.3. Summary

In summary, previous work is limited in several ways. They either require large amounts of training data, are action specific (gait only), cannot be applied to full body motion data, do not preserve the utility of the data, or are simply not effective.

8.3. Methodology

As our Chapter 4 and prior work [251] have shown, anonymizing motion data is a challenging task. The main problem is that motion data contains a large number of dependencies between the individual tracked points, and constraints such as the maximum degree of flexion of certain joints. In addition, the physiology of the person performing the motion is important, as it strongly influences how motions are executed to perform the same action. For example, a tall person will bend their shoulder joint differently to grab an object from a table than a shorter person in the same situation. Because of all these dependencies, directly modifying the data is either not effective because the dependencies can be used to reconstruct the original data, or the modification has to be very strong, which greatly reduces the utility of the motion data. Another interpretation of the dependencies is that they represent redundancy in the data, since the true dimension of the motion data, which can be changed independently, is much smaller than the recorded positions.

The main idea of Pantomime is to remove as many of the dependencies described above as possible before performing anonymization, and then reintroduce the dependencies after anonymization to generate a new sequence of motion data. We perform the dependency removal by mapping a motion sequence into the motion space of a foundation motion model. We then anonymize the motion sequence in the motion space by adding noise to it. Finally, we map the anonymized motion sequence back to its original position space.

8.3.1. Requirements

The two main goals of motion anonymization are to prevent the identification of an individual from their respective motion sequences, and to preserve the utility of the motion sequences for the application for which they are intended. For Pantomime, the utility goals are naturalness and action similarity. Naturalness means that the motion sequence appears as a genuine motion sequence to a human observer. And action similarity means that the action in the anonymized sequence should be as similar as possible to the original one. We chose naturalness because for many applications of full-body motion data, such as social interactions in MR, it is important that the motions appear believable and realistic. The action similarity should prevent our anonymized motion sequences from deviating too much from the original ones, otherwise we could just generate random motion sequences to satisfy the naturalness goal. In addition, it would be beneficial to meet the following requirements derived from common applications of motion data. The first application requirement is the applicability of anonymization to full-body tracking, since even from sparse input data, such as tracking from MR devices, the full-body pose can be estimated. The second is that the anonymization should be general with respect to the format in which the motion data is captured, since motion capture systems vary in the number of points tracked and the specific body landmarks that are captured.

8.3.2. Anonymizing Human Motion

We now explain Pantomimes approach and design. Note that for Pantomime, we only focus on the anonymization of the body pose, the Θ parameter of the SMPL body model. We do not consider the anonymization of the body shape, the root translation, and the root orientation and remove them by setting their respective parameters to zero. We do this because we assume that a person who wants to be anonymous will choose a digital body that does not resemble them, including the body shape. The root translation and orientation can be estimated from the resulting motion sequence.

As a first step, we unify different motion capture formats into the format of the SMPL model by performing a fitting step. This step is necessary because we want Pantomime to be general enough to work with different formats of motion data. After transforming the original data into SMPL data, we map the data into the motion space of a foundation motion model and then perform anonymization by adding noise to the data. The intuition here is that since the motion space encodes plausible motions, by modifying the input data in this space, we end up with a plausible motion for the output of the anonymization. In contrast, performing anonymization by adding noise to the original position data quickly leads to implausible motions because the individual points are modified without adhering to the given physiology of the body or physics. The final step is to decode the anonymized motion space data back into the SMPL model format and then back into the original motion data format.

In the following, we describe the different steps of our anonymization pipeline for Pantomime, an overview of the whole process can be seen in Figure 8.1.

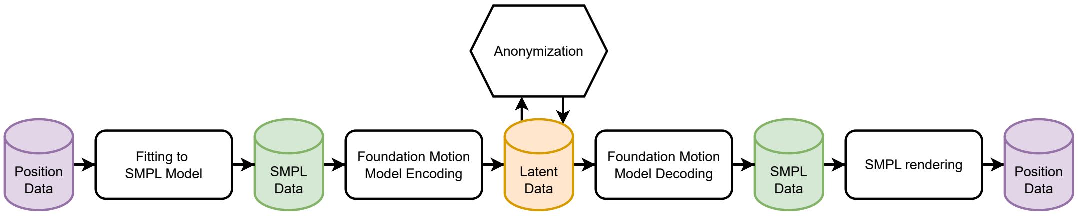


Figure 8.1.: The Pantomime anonymization pipeline.

SMPL Model Fitting

To unify the motion data formats, we perform a fitting of our original motion sequence to the SMPL body model using the HuMoR [306] fitting method. This way all of our motion sequences have the same format and can be used with the same foundation motion models. The fitting process is a function optimization that is performed for each pose of a motion sequence to find the best SMPL parameters to represent that pose.

Given an observed motion sequence, we try to find the parameters of the SMPL body model $M(r, \Phi, \Theta, \beta)$ that best describe the sequence. To perform this function optimization, three loss terms must be minimized. The first is the plausibility of the motion. Here we use either VPoser or HuMoR as a prior to measure the plausibility of a given pose (VPoser) or the plausibility of a given pose transition (HuMoR). The second loss term is a reconstruction loss comparing the original with the found SMPL joint positions using MSE. Since the skeletons between motion data representations may have different kinematic structures, we define a mapping between the SMPL body model skeleton and the dataset we are trying to fit to compute this reconstruction loss. The third loss is a regularization term that enforces consistency of bone length, ground contact, ground position, and body shape across the entire motion sequence (i.e., bone length and body shape should not change much during a single motion sequence). The weighting of the three loss parts is dataset specific and is determined by performing a hyperparameter optimization to find the optimal values. A more detailed description of the loss terms can be found in [306].

This first step can also be seen as a decoupling of the dependencies, since the original position data is split into the body shape β and the joint angles Θ, effectively decoupling these two aspects of the data.

Encoding the Motion Sequences:

Next, we encode the motion sequences into the latent space of a foundation motion model. The foundation model used for this step is interchangeable, as Pantomime only requires that it be a VAE that encodes from the SMPL parameters to a latent space of plausible poses or pose transitions and back to the SMPL parameters. The rationale behind this requirement is that VAEs compress the original data by removing dependencies [337, 43] (i.e., correlations) between data points, and thus our anonymization no longer needs to adhere to these dependencies for effective anonymization. In other words, by performing anonymization in the space of plausible motions, the resulting anonymized motion is itself plausible, thus preserving the utility of the data. In this chapter, we use VPoser and HuMoR as two possible foundation motion models that satisfy this requirement. However, it is important to note that VPoser only takes the body pose (joint angles Θ) as input, while HuMoR also requires the root orientation, root translation, and body shape.

Anonymizing the Latent Code:

Now we anonymize the latent code. To do this, we draw a noise vector p from a normal distribution $\mathcal{N}(\mu, \sigma)$ with $\mu = 0$ and $\sigma = 1$ and add it to the latent code z_t at time step t. We chose normal distributed noise because VAEs regularize their latent space to be normal distributed and adding two normal distributed random variables results in a normal distributed sum [181]. Since the decoder expects a normally distributed random variable as input, this should result in the least utility loss. We chose the mean μ to be 0 and the σ to be 1 because we do not want to introduce a bias into the anonymized motion sequences. With a mean of 0, the resulting distribution of anonymized motion sequences will cluster around the original motion sequence. The noise vector p is scaled by the scalar γ to make the anonymization configurable by increasing the distance in motion space to the original sequence. We expect that, as the value of γ increases, the anonymization will become stronger, and the utility of the data will decrease accordingly. There are two different modes for adding the noise. In *variable* we draw a new noise vector p_t for each time-step z_t, giving us $a_t = \gamma p_t + z_t$ for the anonymized latent code a_t, while in *static* we add the same noise vector p to each z_t, giving us $a_t = \gamma p + z_t$.

Decoding the Motion Sequence:

The last step is to decode the anonymized latent code a_t into the parameters of the SMPL body model using the motion model, and from there back into the original position format.

8.4. Evaluation

We now evaluate the privacy-utility tradeoff of Pantomime to understand how much noise must be added for effective anonymization and how much utility is retained. We also test the assumptions we made when designing Pantomime.

8.4.1. Datasets

We select our datasets to include a large number of full body motion capture sequences with a preference for the gait task, as this has been shown to be highly identifiable. We specifically did not select the AMASS [207] dataset, or any of the datasets included in it, because AMASS was used to train VPoser and HuMoR.

For our evaluation, we use our own *CeTI-Locomotion* dataset (see Chapter 7) and Horst-DB [134] dataset. CeTI-Locomotion includes a variety of walking modalities as well as sit-to-stand exercises in which participants stand up and then sit down as quickly as possible. The dataset was recorded from 50 healthy participants using an IMU suit that captures the relative motion of each body part. In combination with the anthropocentric measurements of the participants, the 17 body segment positions are calculated. Horst-DB contains only one walking modality and was recorded using optical motion tracking with 54 reflective markers attached to body landmarks such as joints or the iliac crest. The resulting data are scalar values for the x,y,z positions of each of the markers. The main difference between the two datasets is that CeTI-Locomotion includes several different walking modalities (normal, fast, carrying a backpack, and carrying a bottle crate) plus an additional sit-to-stand exercise. Also, IMU tracking is less accurate than optical marker tracking, which is considered the gold standard of motion tracking. See Table 8.1 for a comparison of the two datasets.

Table 8.1.: Overview numbers of the used evaluation dataset

Name	Points	Participants	Tasks	Samples
CeTI-Locomotion	17	50	5	4672
Horst-DB	54	57	1	1140

8.4.2. Implementation

Here we describe the implementation details for both Pantomime and the biometric recognition systems we will use for our experiments.

Data preparation

We preprocess both our evaluation datasets to have a frame rate of 30 Hz, same as prior work [306]. For the Horst-DB dataset we also additionally cut the samples to exactly one gait cycle (see Section 4.1.1) using the additional force plate data and a threshold to identify the first and the last pose of the cycle, as described by Horst et al. [134].

SMPL parameter fitting

For the translation from the original data format of the datasets to the SMPL format, we perform a fitting step for each motion sequence. Overall, the goal of the fitting is to find the parameters of the SMPL model ($M(r, \Phi, \Theta, \beta)$) that match the positions of the input data as closely as possible, as well as to achieve a high plausibility with the used foundation motion model (here HuMoR or VPoser). The foundation motion model is used here to enforce the generation of only plausible motions for the SMPL poses. To fit our two evaluation datasets we use the code of HuMoR [306], which implements the whole process in three steps. In the first stage, the root translation and rotation are optimized using VPoser as a prior, in the second stage, the entire SMPL parameters are optimized using VPoser as a prior, and in the third stage, HuMoR is used as a prior for the optimization. Since the code performs the optimization first with VPoser and then with HuMoR, we use it to generate both fits.

To perform the fitting, we need a mapping from our motion data joint positions to the SMPL joint positions (i.e., which positions in our data correspond to which position in the SMPL joints). We created the mapping manually as follows. For CeTI-Locomotion this mapping is undercomplete because CeTI-Locomotion has only 17 joints while the SMPL model has 21. For the Horst DB, it is the other way around because the datasets tracked 54 points, some of which are ignored while others are combined to better match the joint positions of the SMPL body model. Joint positions for which there is no matching position in the original data are set to infinity. To obtain good fitting results, we perform hyperparameter optimization on both datasets to find the weights for the fitting losses described in section 8.3.2. Due to the length of the fitting process, we perform the hyperparameter optimization on 10 random motion sequences of each dataset and then use the found parameters for all of them. For both datasets we had some motion sequences for which the SMPL fitting process failed (producing NaN values at some stage of the fitting process) and we were unable to obtain an SMPL body model representation, these sequences were excluded from the datasets. In both datasets this was less than 1% of the total motion sequences.

As Pantomime removes the root translation from the sequences, we estimate a new root translation for the anonymized sequences by using the absolute trajectory of the right foot as the root translation.

Biometric recognition system

For the biometric recognition system, we adapt a state-of-the-art method used in Chapter 4, in which the motion sequence is resampled to a fixed 100 frames and then flattened into a single vector. The samples are then divided into 80% training data and 20% test data in a stratified fashion, while keeping the same percentage of samples per identity class. We use the same system for action recognition, but there we split the data so that a single participant's samples are either in the test data or in the training data. This should prevent the system from learning actions for specific people and help with generalization. The two datasets are then processed independently. Each dimension of the feature vector is min-max normalized before its dimensionality is reduced using a PCA. As a classifier, we use a SVM with a *radial basis function* (RBF) kernel, which is trained in a 5-fold stratified cross-validation procedure using balanced accuracy as a metric. The final result is obtained by running the SVM on the test data. The recognition system is always trained on the anonymized data of the anonymization we are testing, as defined by our adversary model.

Code Availability

We implemented the biometric recognition system using python, scikit-learn, and PyTorch. Pantomime itself is implemented in python on top of the existing HuMoR code [306].

8.4.3. Experiments

Here we detail the experiments we performed to investigate our underlying assumptions in designing Pantomime, and then to evaluate Pantomime's privacy-utility tradeoff.

Assumptions & Baseline

Baseline Identification & Action Recognition First, we establish an identification and action recognition baseline on the original position data, against which we will later compare the anonymization results of Pantomime. This will allow us to evaluate how good Pantomime's privacy protection is and how much utility we lose as a result. In our baseline identification experiment **E1**, we train and test the biometric recognition system on the dataset to perform person identification. To do this, we split the dataset into a test and a training part of the dataset, with each person having different samples in both partitions. We then train the biometric recognition system on the training dataset in a supervised manner. We then determine the identification performance on the test set. In the action recognition baseline experiment **E2**, we determine the baseline action recognition performance. We now train and test an action recognition system. Unlike for person identification, we split the dataset so that a person is either in the test or training dataset to avoid the system learning the unique action performance of a person and to better generalize. We then measure how well the system can identify the action performed in the sample.

Identification Potential SMPL Parameters The unification of the motion data in the SMPL body format splits the data into motion data (poses as joint angles Θ), body shape β, root translation r, and root orientation Φ. Pantomime focuses only on anonymizing the motion data of the poses. For the remaining SMPL parameters, we test how much identification potential they have on their own.

For **E3** we use the SMPL body model fits from both VPoser and HuMoR for our motion data where specific components are removed by setting them to zero. We then generate new position data for the classification. We expect the body shape and poses alone to carry a high identification potential, as has been shown in previous studies [115, 319]. For the root translation and especially for the root rotation we expect a lower identification potential, because these are single vectors, which should carry less information than the body poses.

Dependency Reduction It is our assumption that by encoding our motion sequences using the SMPL body model and then foundation motion models, the dependencies between the individual data dimensions of a pose or pose transition are reduced due to the VAE architecture used in the foundation models. For **E4** we measure the linear dependence of the pose dimensions on each other. We do this by measuring the average absolute covariance between all dimensions of a pose and then averaging them over the number of poses per motion sequence. This gives us a single comparable linear dependency measure per motion sequence. We then compare the dependency for the different encodings of the motion sequence (original and latent code) to see if the dependency is reduced by the encoding of the foundation motion models.

Noise Mode Comparison We assume that adding the same noise vector to all poses of a motion sequence (static noise mode) will perform better than adding a new noise vector to each pose (variable noise mode). In our noise mode comparison experiment **E5**, we run both modes with different values of the noise scaling γ on the original motion data to test this assumption. We then measure privacy by performing identification with our biometric recognition system. We expect that the static noise mode will always outperform the variable noise mode. Since a motion sequence is a time series of poses, two consecutive poses will be very similar to each other because a person cannot move much in a single time step. Adding different noise vectors to two similar poses makes it easier to separate the noise from the underlying data, since much of the difference between the two poses after the noise is added is the noise. If we add the same noise vector to both poses, then the difference between the two poses is still the same as it was before anonymization, and we cannot distinguish noise from data.

Privacy-Utility Evaluation

In our privacy-utility experiment **E6** we evaluate the privacy and utility of Pantomime. For a better comparison, we not only study how noise injection into the latent space affects the privacy-utility tradeoff, but also test applying noise directly to the original data and the fitted SMPL representation. Due to the different representation of the motion sequence (original, SMPL fit, latent code), the noise parameters of our different anonymization techniques are not directly comparable. For example, adding the same amount of noise to the position of a joint will have a different effect than adding noise to the joint rotations of an SMPL fit. In order to achieve comparability between where we apply noise, we define protection targets. A protection target is a given value of recognition accuracy, for example 20%. We then tune the noise parameters of our anonymization to achieve the given target ($\pm 2\%$). In this way, the anonymization performance of our techniques is the same, and we can directly compare the utility of our approaches to judge which anonymization has the better tradeoff. An overview of all combinations of motion representation, anonymization, and protection target is given in Table 8.2.

Table 8.2.: The combinations of motion representations, anonymizations, and protection targets which we use in our privacy-utility tradeoff.

motion rep.	anonymization	protection targets
original	direct	10%, 20%
SMPL (VPoser)	direct, VPoser, HuMoR	10%, 20%
SMPL (HuMoR)	direct, VPoser, HuMoR	10%, 20%

We investigate utility by performing action recognition on the CeTI-Locomotion dataset and by conducting a user study on both datasets. We investigate two utility goals in our user study, the first is the naturalness of an anonymized motion sequence, and the second is the motion similarity between an original and corresponding anonymized motion sequence.

VPoser only works on a single pose at a time, while HuMoR works on the pose transition and thus on pose pairs. Due to this, and the better overall performance for pose fitting reported by HuMoR [306], we expect HuMoR to do a better job of removing the dependencies, and therefore Pantomime to achieve better utility. Furthermore, we expect Pantomime to achieve better utility than adding noise to the original pose data, regardless of the foundation motion model chosen.

8.4.4. Utility / User Study

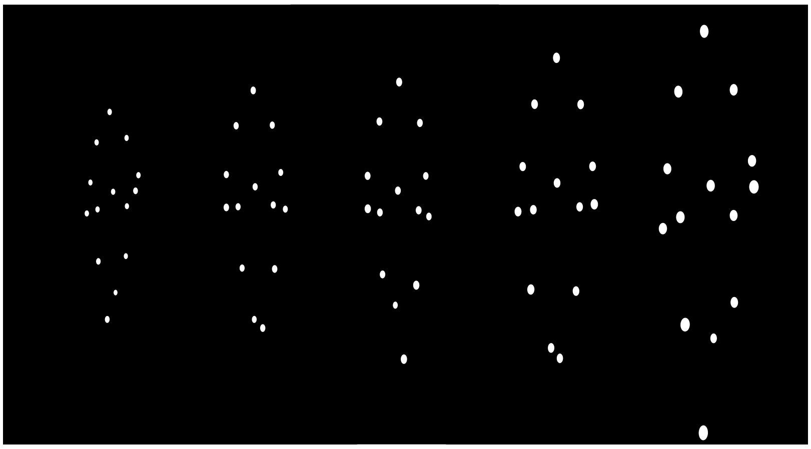


Figure 8.2.: Example rendering of a motion sequence used during the user study.

For utility, we investigate how natural the anonymized motion sequences appear and how similar the actions are compared to the original motion sequence. We do this by conducting a user study similar to another study [355] we performed on the evaluation of biometric anonymizations. We use two tasks to evaluate our objectives. In the first task, we show participants a single motion sequence. The participants then rate how natural the sequences appear by answering the question "Is this a natural human motion? The rating is done on a 5-point Likert scale from "very unnatural" (1) to "very natural" (5). In the second task, we show the original sequence next to an anonymized version of the sequence. Participants then answer the question "How similar are the motions performed in the two videos?" by rating on a 5-point Likert scale from "very dissimilar" (1) to "very similar" (5).

We conducted an online survey in which participants were shown different motion sequences from the CeTI-Locomotion and Horst-DB datasets as short video sequences (see Figure 8.2 for an example) rendered at 20 frames per second. We reduced the rendering from 30 to 20 frames per second to allow users to better judge the execution of the motion. The motion is rendered as point-light displays to reduce the influence of appearance and to focus on the motion. From the Horst DB, we randomly select 4 motion sequences, two from male participants and two from female participants. From CeTI-Locomotion, we randomly select two male and two female motion sequences for the modalities *gait-normal*, *gait-fast*, *gait-bottle-crate*, and *sit-to-stand*. These 20 motion sequences were anonymized with a subset (original+direct, HuMoR+direct, HuMoR+HuMoR, VPoser+direct, and VPoser+VPoser) of the combinations described above (see Table 8.2) for the protection targets of 10% and 20%. Since we perform two tasks (naturalness and similarity see above) and add the original sequences, this results in 440 unique questions. From this question pool, each participant answers 40 random questions.

8.4.5. Ethical Considerations

The user study data collection was approved by the ethics commission of the Karlsruhe Institute of Technology (research project "Utility of Anonymized Motion Sequences") and was conducted in accordance with the Declaration of Helsinki. All data was collected in an anonymous online survey in December 2024 using an online recruitment platform [1] to recruit 224 participants (112 male, 112 female; mean age 30.8, std 9.57). Participation took a median of 7:47 minutes and participants were paid an average of 10.71£ per hour.

The CeTI-Locomotion and Horst-DB dataset used in this study both had approval by their respective ethics commissions and their participants gave informed written consent to participate in the data collection.

8.5. Results

Here, we describe the results of our experiments, similar to the evaluation section, where we start with the results of testing the underlying assumptions for Pantomime before describing the results of the privacy-utility evaluation.

8.5.1. Assumptions & Baseline

We start with the baseline identification experiment **E1** and the action recognition experiment **E2**. For both CeTI-Locomotion and Horst-DB we observe that the identification recognition accuracy for the original data is high with 83% and 96%, respectively. For the CeTI-Locomotion dataset, we also perform action recognition to classify which of the 5 actions was performed in a motion sequence, achieving an accuracy of 80%. Overall, these results are in line with prior work and show that the recognition systems used for identification and action recognition work.

For **E3** we look at the difference in identification accuracy when using only certain parts of the data in their SMPL representation. In Table 8.3 we report the identification accuracy for the data when using positions generated from the SMPL representation with only the

[1] https://prolific.com

Table 8.3.: The influence of the individual parameters of the VPoser and Humor SMPL fits on the identification given as accuracy

datasets SMPL para.		CeTI-Locomotion	Horst-DB
VPoser	shape (β)	0.52	0.75
	joint poses (Θ)	0.63	0.69
	root trans. (r)	0.65	1.0
	root orient. (Φ)	0.61	0.75
HuMoR	shape (β)	0.38	0.6
	joint poses (Θ)	0.2	0.41
	root trans. (r)	0.21	0.64
	root orient. (Φ)	0.31	0.55

Table 8.4.: Average absolute linear correlation between all of the dimensions of the respective encoding of a motion sequence.

	CeTI-Locomotion		Horst-DB	
SMPL Fit	VPoser	HuMoR	VPoser	HuMoR
positions	0.55	0.62	0.46	0.64
joint poses	0.55	0.55	0.55	0.5
VPoser lat. enc.	0.5	0.52	0.41	0.46
HuMoR lat enc.	0.52	0.37	0.34	0.39

specific parameter intact, while the rest of the parameters are set to zero. We find that most parameters have a high identification potential (greater than 50%) on their own. The lowest identification potential is observed for the SMPL fit of the CeTI-Locomotion data using HuMoR. But even here, the individual parameters have significantly more identification potential than the chance level (2%) for CeTI-Locomotion. This leads us to conclude that anonymizing only the SMPL joint poses in the latent space is not sufficient, as the remaining components of the SMPL body model can be used for identification. This justifies our decision to set the remaining parameters to zero.

For **E4** we hypothesized that using the foundation motion models will reduce the dependency between the individual data dimensions. In Table 8.4 we report the average absolute correlation between the different data dimensions for different representations of the motion sequences. We find that the data represented as positions (generated from the respective SMPL fit) has the highest average absolute correlation, with the exception of the Horst-DB VPoser fit. The data represented as joint poses has the second highest, followed by the latent encodings of VPoser and HuMoR. This decrease in correlation is in line with our expectations, and especially HuMoR seems to achieve a high decoupling of the latent dimensions. However, the remaining high correlations for the VPoser latent encoding show that the effects of decoupling can be much smaller than expected.

Next we report the results for our noise mode experiment **E5**, see Figure 8.3. For both datasets it can be seen that when using the same noise scaling γ the variable noise mode always is outperformed by the static noise mode. When visually inspecting the resulting motion sequences using a rendering of the sequence it can also be seen that the approach

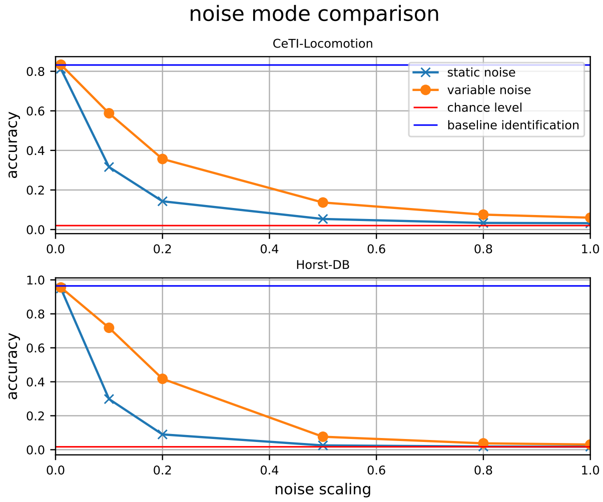


Figure 8.3.: Comparison of the static noise vs. variable noise mode of directly applying noise to the original data of both CeTI-Locomotion and Horst-DB over different γ noise scaling choices.

leads to a very visible shaking of the joint points, the injected noise becomes visible in the motion execution.

Since two consecutive poses are very similar to each other, changing the noise vector for each pose makes it easier to distinguish what is the real data and what is the noise addition. Our conclusion from this experiment is that the static noise mode is the better mode to add the noise to the data as it always outperforms the variable noise for the identification reduction.

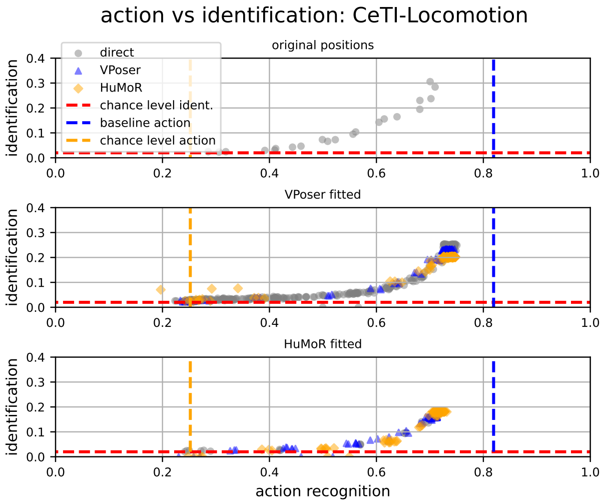


Figure 8.4.: The action recognition vs identity recognition accuracy for CeTI-Locomotion.

To investigate the privacy-utility tradeoff **E6**, we first report the action recognition results for the CeTI-Locomotion dataset, see Figure 8.4, before doing the main comparison using our user study. Comparing the direct anonymization on the original data with all the anonymizations on the adjusted data, we see that the adjusted data drops faster in person identification than in action recognition. For example, at 60% action recognition accuracy, the original data is still at about 15% identification accuracy, while the fitted data is much lower and close to 5% accuracy. Surprisingly, there does not seem to be a difference if the anonymization on the fitted data is performed directly on the positions, the SMPL joints, or the latent encoding, as all these anonymizations perform similarly. At least for action recognition, it does not seem to matter in which data space the anonymization is performed. We conclude that action recognition is a simple task that is still successful on heavily distorted data, which leads to similar performance of the anonymizations.

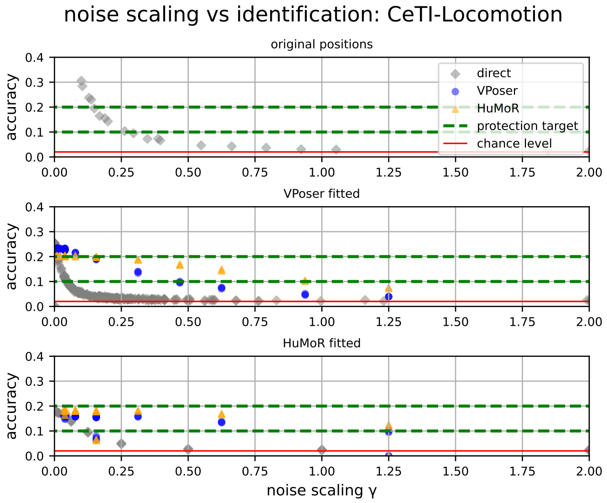


Figure 8.5.: The scaling of the noise parameter γ in comparison to the achieved recognition accuracy of different anonymization combinations for CeTI-Locomotion.

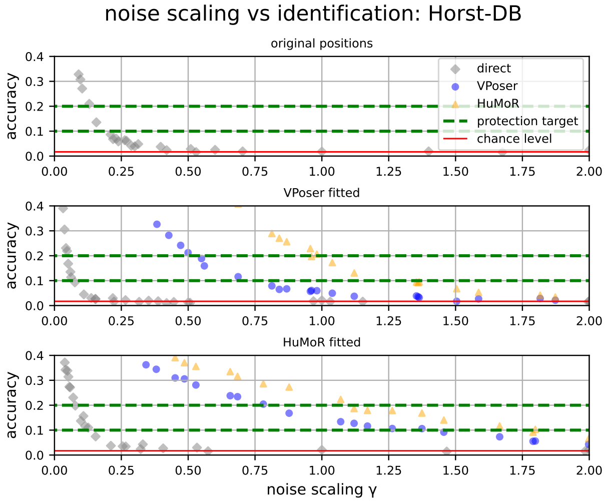


Figure 8.6.: The scaling of the noise parameter γ in comparison to the achieved recognition accuracy of different anonymization combinations for Horst-DB.

For the main privacy-utility tradeoff, we first report the scaling of the noise parameter versus the achieved recognition accuracy, see Figure 8.5 and Figure 8.6. We use these results to pick the noise scaling values which fulfill the specific protection targets (shown as green line). We find that the direct application of noise to the position data requires less noise scaling than when we apply the noise in the latent space or directly to the joint rotations of the SMPL model. For both datasets, we choose the anonymizations that achieve 10% or 20% ($\pm 2\%$) for the direct comparison of utility in the user study.

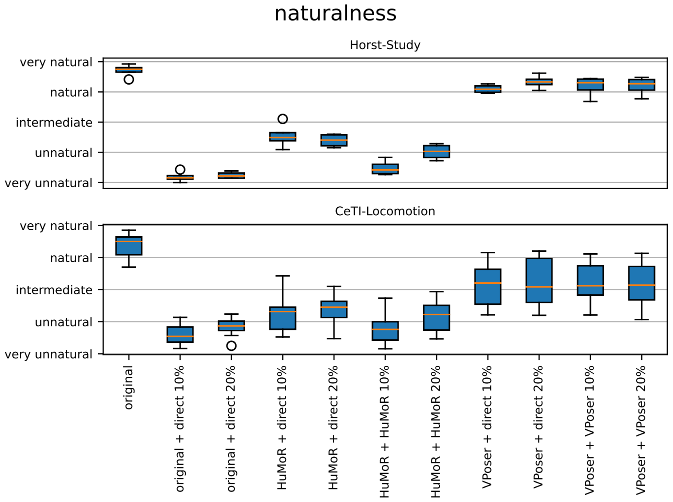


Figure 8.7.: Average naturalness rating per question of the user study as a box plot per anonymization.

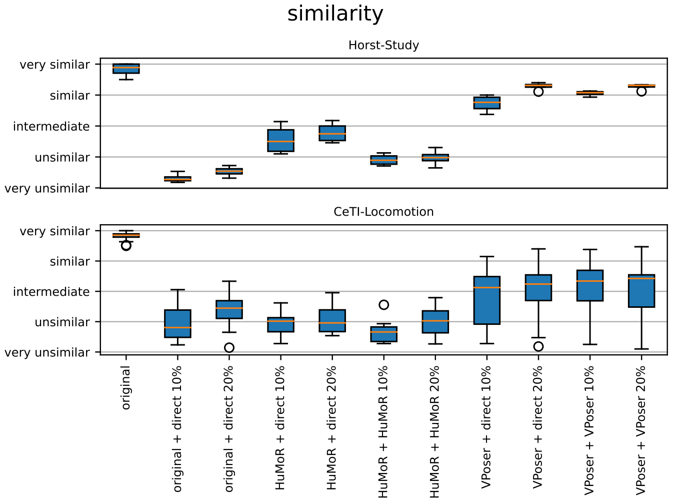


Figure 8.8.: Average similarity rating per question of the user study as a box plot per anonymization.

For our user study, we first calculate the average of the ratings for each question, and then use the averages to create a box plot for each anonymization technique. Comparing the naturalness results (see Figure 8.7) we see that the original motion sequences score the highest for both datasets, indicating that the participants consider the selected datasets to be representative of natural motion. For the direct addition of noise to the original position data, we see that both datasets have very low naturalness ratings, close to very unnatural, showing that the naturalness of the motion data is destroyed. The anonymizations using HuMoR as a foundation model achieve more utility than the direct anonymization, but also all are rated close to unnatural. The best natural ratings we see for the anonymization using VPoser, for the CeTI-Locomotion dataset an intermediate rating is achieved, while for the Horst-DB the rating is even above natural. We conclude that using VPoser as a foundation model during anonymization helps to achieve natural motion sequences that are close to the naturalness of the original.

For the similarity results (see Figure 8.8), we generally see a similar pattern as for naturalness, with the original data rated as very similar to itself. Direct anonymization on the original data achieves almost no similarity, HuMoR achieves intermediate to unsimilar results, and VPoser again delivers the best similarity on the Horst-DB datasets. This again shows that using VPoser, Pantomime can successfully anonymize while keeping the motion sequence similar to the original, thus preserving utility.

8.5.2. Summary of Results

Here, we give a short summary of the main results of our evaluation:

- Pantomime is able to successfully anonymize motion data by anonymizing it in the latent space of a foundation motion model.
- All components of the SMPL representation of the motion sequences contain identifiable information.
- The latent encodings of motion sequences using foundation motion models only slightly reduce the correlations between the data dimensions.
- Applying a fixed random vector to the entire motion sequence instead of varying it for each pose is the better mode for anonymizing motion sequences.
- For the action recognition we do not see a significant difference between the anonymization techniques.
- Using a VPoser fitting with a VPoser latent encoding achieves the best privacy-utility tradeoff.

8.6. Discussion, Limitations & Future Work

In general, we find that Pantomime's approach of using foundation motion models to first fit the position data to the SMPL body model and then anonymize the data in the latent space of the model is a viable approach to anonymize full-body motion data in a plausible manner. Because of the plausibility constraints added by the foundation models, the data can retain the utility of the motion data while performing effective anonymization.

In comparison with prior work, Pantomime has some key advantages. It anonymizes full-body motion capture data, it is not designed to work only with specific motions (such as gait), it does not require a large corpus of specific application data to train, and it is configurable via its noise scaling, allowing its privacy-utility tradeoff to be adjusted for specific applications. Furthermore, it is general as it performs a unification of motion data formats by fitting to the SMPL body model.

The same foundation models that we use to anonymize can also be used to generate synthetic motion data, by using the original data as an anchor in the latent space and then shifting it by adding noise, we essentially generate synthetic data that is similar to the original data. Thus, Pantomime can be considered as synthetic data generation. Similar to face anonymization [140], which generates new faces to anonymize facial images that have similar characteristics such as ethnicity or age.

Pantomime also has some limitations that need to be addressed. Its main drawback is the poor fitting quality for some of the motion sequences, especially when HuMoR is used for fitting. Another problem with fitting is that it does not work well for certain actions, such as the stand-to-sit tasks in CeTI-Locomotion. Furthermore, the current implementation of the fitting process is slow (it takes about 1.5 weeks to process CeTI-Locomotion on a single GeForce RTX 3090). However, the anonymization process itself is much faster. It takes only about 16 minutes to anonymize the entire CeTI-Locomotion dataset, which contains about 1.5 hours of recorded sequences.

Pantomime's approach to anonymization is very general and could be promising for other complex data types that are difficult to anonymize, such as trajectories or other biometric features. The only requirement would be that the foundation models use a VAE structure. Furthermore, it should be investigated how the approach can be realized in a faster fashion to allow for applications like the real-time streaming of motion data.

8.7. Chapter Summary

We have presented Pantomime, a full-body motion anonymization that uses foundation motion models to anonymize in a plausible way. Different from prior work Pantomime does not require the training of a specialized anonymization model for a specific dataset but uses existing foundation motion models. Another key advantage of Pantomime is that its anonymization strength can be configured via its noise scaling. Our results show that using the VPoser model it is possible to achieve identification rates as low as 10% while keeping the anonymized motions natural and similar to the original ones. This is an important step towards a more privacy-preserving use of motion tracking in applications such as MR, robotics, or medicine.

9. Conclusion

Human motion data is a rich source of information with a wide range of applications. In medicine, for example, it can be used to diagnose diseases and monitor patient recovery. In robotics, motion data can be used to teach robots new skills. In MR, captured motion data allows for the animation of digital avatars, creating better social interaction between users. As these application fields continue to develop, the use of motion capture technology will likely become a normal part of everyday life.

However, motion data has the downside of being a behavioral biometric trait, which poses a privacy risk to the person being recorded. Individuals can be identified by their gait or how they use MR headsets. Furthermore, private attributes can be inferred, such as sex, age, and medical status. This thesis focused on the privacy risks and mitigations associated with the usage of motion data.

We started by conducting the first literature survey on the anonymization of behavioral biometric data, comparing anonymization techniques across multiple behavioral traits. For the classification of the found techniques we developed a new general taxonomy which enables the comparison of different anonymization independent of their behavioral trait. This helps with identifying suitable anonymizations for one trait that can be adapted for another one, in our case motion data. Further, our survey found that the largest corpus of literature focus on anonymizations for voice recordings. Unsurprisingly the state-of-the-art of voice anonymization is much more advanced then for other traits. One example of this is the VoicePrivacy challenge, a community benchmark that seeks to establish a common protocol and datasets for the evaluation of voice anonymizations. Many other behavioral biometric traits would benefit from having such a benchmark, as most performance results are not comparable between different studies. Additionally, our review of evaluation methodologies showed that they are similar across traits and that they suffer from the same unrealistic assumption about the attacker, and hence the reported anonymization protection is likely not reliable. We addressed this by proposing suitable improvements to the evaluation of anonymizations for behavioral biometric data in this thesis.

We investigated which features of gait motion data contribute to identifying individuals and recognizing sex. Our main finding was that identification and sex recognition are both very robust, and no simple perturbation can prevent this without drastically decreasing the data's utility. One possible interpretation of this finding is that gait data is idiosyncratic and redundant. This suggests that effectively anonymizing gait data is a hard problem and requires an approach that considers the interdependence of gait data. Modifying a single feature or data point without adjusting the others accordingly will produce unrealistic results and weaken anonymization, as modifications can be spotted and distinguished from real data. While this effect is strong with motion data, we expect to see similar effects with other complex data types, where all data points are interconnected.

To further study the privacy risks of motion data, we collected the first comprehensive dataset of facial motions using MR headsets. Using this data, we demonstrated that facial motion can be used to identify individuals with up to 98% accuracy in a single session. For identification across multiple sessions, accuracy is much lower, but still much higher than chance level at 43%. Since we used different types of MR headsets, we also demonstrated that identification is possible across different headset types. This highlights that switching to a new headset type will not protect users' privacy. Overall, we conclude that facial motion data poses a significant privacy risk to the person who was recorded and should be treated as personally identifiable information.

We proposed improvements to the methodology for evaluating biometric data anonymization to address the major flaws that we identified in our survey. We demonstrated the importance of retraining the biometric recognition model, as this allows the model to adapt to the anonymized data. This results in more reliable outcomes when the assumed attacker is stronger. Furthermore, we demonstrated the importance of using different biometric systems because the system that performs best with clear data is not necessarily the best with anonymized data. Finally, we demonstrated that decreasing the number of identities in the evaluation dataset and carefully selecting which identities to include simulates a worst-case anonymization scenario. These improvements to the evaluation methodology result in a more rigorous evaluation of anonymization performance and more reliable results.

To study the anonymization of motion data further, we collected the CeTI-Locomotion dataset. The dataset was collected from 50 participants performing five different motion tasks while wearing motion capture suits. Due to the large number of repetitions for each task, this dataset is ideal for identifying individuals and testing anonymization approaches.

Lastly, we proposed Pantomime, the first anonymization technique for full-body motion data that enables high utility while protecting against re-identification. Pantomime first moves the motion data into the latent space of foundation motion models to anonymize it before adding noise. Due to the addition of noise in the latent space, the resulting anonymized data is a valid motion sequence similar to the original. Our user study showed that the anonymized data achieves high naturalness and similarity. This makes Pantomime the first anonymization method that can effectively anonymize motion data while maintaining high utility.

Open Challenges Although we addressed important questions regarding the anonymization of motion data, some questions remain unanswered.

The most pressing issue we currently face is the applicability of motion anonymization, including Pantomime, to real-world scenarios. Many applications of motion data would benefit from real-time anonymization capabilities. This would allow for the anonymization of motion data streams. For example, it could be used for the real-time animation of digital avatars in MR. In principle, Pantomime can process streams because it anonymizes one pose at a time. However, its current implementation uses a slow fitting process that prevents this use case.

Furthermore, motion data should be anonymized as early as possible to reduce the potential attack surface. Therefore, the best place to anonymize motion data is on the recording device. However, Pantomime and other machine learning approaches require dedicated GPUs and much processing power. For this reason, we need more lightweight approaches that can be deployed on mobile devices, such as MR headsets.

As we have demonstrated in this thesis, facial motion data is a behavioral biometric trait that can be used for identification purposes. Currently, there is no anonymization method for facial motion data. While Pantomime can be adopted for this task in principle, there are currently no foundation models for facial motion that could be used for anonymization.

In this thesis, we examined motion data captured using 3D motion capture technology. However, most motion data is still recorded in the form of 2D videos. Anonymizing this type of data comes with the additional challenge of producing visually believable results that fit well into the rest of the video. Some early works address this problem; however, many open questions remain.

Bibliography

[1] Alberto Abad, Alfonso Ortega, António Teixeira, Carmen García Mateo, Carlos D. Martínez Hinarejos, Fernando Perdigão, Fernando Batista, and Nuno Mamede. *Advances in Speech and Language Technologies for Iberian Languages*. 2016, DOI: 10.1007/978-3-319-49169-1.

[2] Enas Abdulhay, N Arunkumar, Kumaravelu Narasimhan, Elamaran Vellaiappan, and V Venkatraman. "Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease". In: *Future Generation Computer Systems*, 2018, pp. 366–373, DOI: 10.1016/j.future.2018.02.009.

[3] Mohammed Abo-Zahhad, Sabah Mohammed Ahmed, and Sherif Nagib Abbas. "State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals". In: *Biometrics*, 2015, pp. 179–190, DOI: 10.1049/iet-bmt.2014.0040.

[4] Mohamed Abou-Zleikha, Zheng-Hua Tan, Mads Graesboll Christensen, and Soren Holdt Jensen. "A discriminative approach for speaker selection in speaker de-identification systems". In: *European Signal Processing Conference*, 2015, pp. 2102–2106, DOI: 10.1109/eusipco.2015.7362755.

[5] Richard A. Abrams, David E. Meyer, and Sylvan Kornblum. "Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system." In: *Journal of Experimental Psychology: Human Perception and Performance*, 1989, pp. 529–543, DOI: 10.1037/0096-1523.15.3.529.

[6] Christopher Ackad, Andrew Clayphan, Roberto Martinez Maldonado, and Judy Kay. "Seamless and continuous user identification for interactive tabletops using personal device handshaking and body tracking". In: *Extended Abstracts on Human Factors in Computing Systems*, 2012, pp. 1775–1780, DOI: 10.1145/2212776.2223708.

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

[8] Ayush Agarwal, Amitabh Swain, and S. R. Mahadeva Prasanna. "Speaker Anonymization for Machines using Sinusoidal Model". In: *IEEE International Conference on Signal Processing and Communications (SPCOM)*, 2022, pp. 1–5, DOI: 10.1109/SPCOM55316.2022.9840792.

[9] Prachi Agrawal and P. J. Narayanan. "Person De-Identification in Videos". In: *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, pp. 299–310, DOI: 10.1109/tcsvt.2011.2105551.

[10] Hafiz Shehbaz Ali, Fakhar ul Hassan, Siddique Latif, Habib Ullah Manzoor, and Junaid Qadir. “Privacy Enhanced Speech Emotion Communication using Deep Learning Aided Edge Computing”. In: *International Conference on Communications Workshops*, 2021, pp. 1–5, DOI: 10.1109/ICCWorkshops50388.2021.9473669.

[11] Abdulaziz Almehmadi and Khalil El-Khatib. “The state of the art in electroencephalogram and access control”. In: *Conference on Communications and Information Technology (ICCIT)*, 2013, pp. 49–54, DOI: 10.1109/iccitechnology.2013.6579521.

[12] Ranya Aloufi, Hamed Haddadi, and David Boyle. “Privacy-preserving Voice Analysis via Disentangled Representations”. In: *Conference on Cloud Computing Security Workshop*, 2020, pp. 1–14, DOI: 10.1145/3411495.3421355.

[13] Arwa Alsultan and Kevin Warwick. “Keystroke dynamics authentication: a survey of free-text methods”. In: *International Journal of Computer Science Issues (IJCSI)*, 2013, p. 1.

[14] Abdulaziz Alzubaidi and Jugal Kalita. “Authentication of Smartphone Users Using Behavioral Biometrics”. In: *Communications Surveys & Tutorials*, 2016, pp. 1998–2026, DOI: 10.1109/comst.2016.2537748.

[15] Marianna Amboni, Carlo Ricciardi, Sofia Cuoco, Leandro Donisi, Antonio Volzone, Gianluca Ricciardelli, Maria Teresa Pellecchia, Gabriella Santangelo, Mario Cesarelli, and Paolo Barone. “Mild Cognitive Impairment Subtypes Are Associated With Peculiar Gait Patterns in Parkinson’s Disease”. In: *Front. Aging Neurosci.*, 2022, p. 781480, DOI: 10.3389/fnagi.2022.781480.

[16] Miguel E Andrés, Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. “Geo-indistinguishability: Differential privacy for location-based systems”. In: *ACM CCS*, 2013, pp. 901–914.

[17] *Apple Vision Pro*. https://www.apple.com/apple-vision-pro/, Accessed: 2025-03-03.

[18] Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. “Saccade landing position prediction for gaze-contingent rendering”. In: *ACM Transactions on Graphics*, 2017, pp. 1–12, DOI: 10.1145/3072959.3073642.

[19] Patricia Arias-Cabarcos, Thilo Habrich, Karen Becker, Christian Becker, and Thorsten Strufe. “Inexpensive brainwave authentication: new techniques and insights on user acceptance”. In: *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 55–72.

[20] Sarker Monojit Asish, Arun K. Kulshreshth, and Christoph W. Borst. “User Identification Utilizing Minimal Eye-Gaze Features in Virtual Reality Applications”. In: *Virtual Worlds*, 2022, pp. 42–61, DOI: 10.3390/virtualworlds1010004.

[21] Tom B"ackström, Okko R"as"anen, Abraham Zewoudie, and Pablo Pérez Zarazaga. *Introduction to Speech Processing*. WebPage, Accessed: 02.02.2021.

[22] A.Terry Bahill, Michael R. Clark, and Lawrence Stark. “The main sequence, a tool for studying human eye movements”. In: *Math. Biosci.*, 1975, pp. 191–204, DOI: 10.1016/0025-5564(75)90075-9.

[23] Fahimeh Bahmaninezhad, Chunlei Zhang, and John Hansen. “Convolutional Neural Network Based Speaker De-Identification”. In: *The Speaker and Language Recognition Workshop*, 2018, pp. 255–260, DOI: 10.21437/odyssey.2018-36.

[24] Max Bain, Jaesung Huh, Tengda Han, and Andrew Zisserman. “WhisperX: Time-Accurate Speech Transcription of Long-Form Audio”. In: *INTERSPEECH 2023*, 2023, DOI: 10.21437/interspeech.2023-78.

[25] Michal Balazia and Petr Sojka. “Gait Recognition from Motion Capture Data”. In: *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018, pp. 1–18, DOI: 10.1145/3152124.

[26] Michal Balazia and Petr Sojka. “You are how you walk: Uncooperative MoCap gait identification for video surveillance with incomplete and noisy data”. In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 208–215, DOI: 10.1109/BTAS.2017.8272700.

[27] JA Beintema and Markus Lappe. “Perception of biological motion without local image motion”. In: *Proceedings of the National Academy of Sciences*, 2002, pp. 5661–5663, DOI: 10.1073/pnas.082483699.

[28] Emmanuelle Bellot, Etienne Abassi, and Liuba Papeo. “Moving Toward versus Away from Another: How Body Motion Direction Changes the Representation of Bodies and Actions in the Visual Cortex”. In: *Cerebral Cortex*, 2021, DOI: 10.1093/cercor/bhaa382.

[29] Lanthao Benedikt, Darren Cosker, Paul L Rosin, and David Marshall. “Assessing the Uniqueness and Permanence of Facial Actions for Use in Biometric Applications”. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 2010, pp. 449–460, DOI: 10.1109/tsmca.2010.2041656.

[30] Zineb Bennis and Pierre-Antoine Gourraud. “Application of a novel Anonymization Method for Electrocardiogram data”. In: *International Conference on Arab Women in Computing*, 2021, pp. 1–5, DOI: 10.1145/3485557.3485581.

[31] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. “Detecting Personality Traits Using Eye-Tracking Data”. In: *Conference on Human Factors in Computing Systems CHI*, 2019, pp. 1–12, DOI: 10.1145/3290605.3300451.

[32] David Bethge, Philipp Hallgarten, Tobias Grosse-Puppendahl, Mohamed Kari, Ralf Mikut, Albrecht Schmidt, and Ozan Ozdenizci. “Domain-Invariant Representation Learning from EEG with Private Encoders”. In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 1236–1240, DOI: 10.1109/ICASSP43922.2022.9747398.

[33] G. Bienvenu and L. Kopp. “Adaptivity to background noise spatial coherence for high resolution passive methods”. In: *ICASSP*, 1980, pp. 307–310, DOI: 10.1109/icassp.1980.1171029.

[34] Stefan Billeb, Christian Rathgeb, Herbert Reininger, Klaus Kasper, and Christoph Busch. “Biometric template protection for speaker recognition based on universal background models”. In: *IET Biometrics*, 2015, pp. 116–126, DOI: 10.1049/iet-bmt.2014.0031.

[35] Jutta Billino and Karin S Pilz. “Motion perception as a model for perceptual aging”. In: *Journal of vision*, 2019, pp. 3–3, DOI: 10.1167/19.4.3.

[36] Danielle Z Bolling, Kevin A Pelphrey, and Martha D Kaiser. “Social inclusion enhances biological motion processing: a functional near-infrared spectroscopy study”. In: *Brain topography*, 2013, pp. 315–325, DOI: 10.1007/s10548-012-0253-y.

[37] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. “On the opportunities and risks of foundation models”. In: *arXiv preprint arXiv:2108.07258*, 2021, DOI: https://doi.org/10.48550/arXiv.2108.07258.

[38] Jean-François Bonastre, Héctor Delgado, Nicholas Evans, Tomi Kinnunen, Kong Aik Lee, Xuechen Liu, Andreas Nautsch, Paul-Gauthier Noé, Jose Patino, Md Sahidullah, Brij Srivastava, Massimiliano Todisco, Natalia Tomashenko, Emmanuel Vincent, Xin Wang, and Junichi Yamagishi. “Benchmarking and challenges in security and privacy for voice biometrics”. In: *2021 ISCA Symposium on Security and Privacy in Speech Communication*, 2021, pp. 52–56, DOI: 10.21437/SPSC.2021-11.

[39] Zillah Boraston and Sarah-Jayne Blakemore. “The application of eye-tracking technology in the study of autism”. In: *Physiol. J.*, 2007, pp. 893–898, DOI: 10.1113/jphysiol.2007.133587.

[40] Efe Bozkir, Onur Günlü, Wolfgang Fuhl, Rafael F. Schaefer, and Enkelejda Kasneci. “Differential Privacy for Eye Tracking with Temporal Correlations”. In: *PLoS ONE*, 2021, e0255979, DOI: 10.1371/journal.pone.0255979.

[41] Talia Brandman and Galit Yovel. “Bodies are represented as wholes rather than their sum of parts in the occipital-temporal cortex”. In: *Cerebral Cortex*, 2016, pp. 530–543, DOI: 10.1093/cercor/bhu205.

[42] Michael Brennan, Sadia Afroz, and Rachel Greenstadt. “Adversarial stylometry”. In: *ACM Transactions on Information and System Security*, 2012, pp. 1–22, DOI: 10.1145/2382448.2382450.

[43] Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. “Understanding disentangling in β-VAE”. In: *arXiv preprint arXiv:1804.03599*, 2018.

[44] W.M. Campbell, D.E Sturim, and D.A. Reynolds. “Support vector machines using GMM supervectors for speaker verification”. In: *IEEE Signal Processing Letters*, 2006, pp. 308–311, DOI: 10.1109/lsp.2006.870086.

[45] Emmanuel J. Candes, Justin Romberg, and Terence Tao. “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information”. In: *Trans. Inf. Theory*, 2006, pp. 489–509, DOI: 10.1109/tit.2005.862083.

[46] Anne M.P. Canuto, Fernando Pintro, and Michael C. Fairhurst. “An effective template protection method for face and voice cancellable identification”. In: *International Journal of Hybrid Intelligent Systems*, 2014, pp. 157–166, DOI: 10.3233/his-140192.

[47] Hyung-Pil Chang, In-Chul Yoo, Changhyeon Jeong, and Dongsuk Yook. “Zero-Shot Unseen Speaker Anonymization via Voice Conversion”. In: *IEEE Access*, 2022, pp. 130190–130199, DOI: 10.1109/ACCESS.2022.3227963.

[48] Meng Chen, Li Lu, Junhao Wang, Jiadi Yu, Yingying Chen, Zhibo Wang, Zhongjie Ba, Feng Lin, and Kui Ren. “VoiceCloak: Adversarial Example Enabled Voice De-Identification with Balanced Privacy and Utility”. In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2023, DOI: 10.1145/3596266.

[49] Peng-Tzu Chen, Shun-Chi Wu, and Jui-Hsuan Hsieh. “A cancelable biometric scheme based on multi-lead ECGs”. In: *IEEE Engineering in Medicine and Biology Society*, 2017, pp. 3497–3500, DOI: 10.1109/embc.2017.8037610.

[50] Xinrun Chen and Hengxin Chen. “Emotion recognition using facial expressions in an immersive virtual reality application”. In: *Virtual Reality*, 2022, pp. 1717–1732, DOI: 10.1007/s10055-022-00720-9.

[51] Yuedong Chen, Jianfeng Wang, Shikai Chen, Zhongchao Shi, and Jianfei Cai. “Facial Motion Prior Networks for Facial Expression Recognition”. In: *2019 IEEE Visual Communications and Image Processing (VCIP)*, 2019, DOI: 10.1109/vcip47243.2019.8965826.

[52] Ming Cheng, Xingjian Diao, Shitong Cheng, and Wenjun Liu. “Saic: Integration of speech anonymization and identity classification”. In: *AI for Health Equity and Fairness: Leveraging AI to Address Social Determinants of Health*. Springer, 2024, pp. 295–306.

[53] Peng Cheng and Utz Roedig. “Personal Voice Assistant Security and Privacy—A Survey”. In: *Proceedings of the IEEE*, 2022, pp. 476–507, DOI: 10.1109/jproc.2022.3153167.

[54] Roman Chereshnev and Attila Kertész-Farkas. “HuGaDB: Human Gait Database for Activity Recognition from Wearable Inertial Sensor Networks”. In: *Analysis of Images, Social Networks and Texts: 6th International Conference, AIST 2017, Moscow, Russia, July 27–29, 2017, Revised Selected Papers 6*, 2018, pp. 131–141, DOI: 10.1007/978-3-319-73013-4_12.

[55] Ching-Yao Chou, En-Jui Chang, Huai-Ting Li, and An-Yeu Wu. “Low-Complexity Privacy-Preserving Compressive Analysis Using Subspace-Based Dictionary for ECG Telemonitoring System”. In: *IEEE Transactions on Biomedical Circuits and Systems*, 2018, pp. 801–811, DOI: 10.1109/tbcas.2018.2828031.

[56] Oubaïda Chouchane, Michele Panariello, Chiara Galdi, Massimiliano Todisco, and Nicholas Evans. “Fairness and Privacy in Voice Biometrics: A Study of Gender Influences Using wav2vec 2.0”. In: *2023 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2023, pp. 1–7, DOI: 10.1109/BIOSIG58226.2023.10345975.

[57] Alice Cohen-Hadria, Mark Cartwright, Brian McFee, and Juan Pablo Bello. “Voice Anonymization in Urban Sound Recordings”. In: *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2019, pp. 1–6, DOI: 10.1109/MLSP.2019.8918913.

[58] Cristina Conati, Christina Merten, Saleema Amershi, and Kasia Muldner. “Using eye-tracking data for high-level user modeling in adaptive interfaces”. In: *AAAI*, 2007, pp. 1614–1617.

[59] Patrick Connor and Arun Ross. “Biometric Recognition by Gait: A Survey of Modalities and Features”. In: *Computer Vision and Image Understanding*, 2018, pp. 1–27, DOI: 10.1016/j.cviu.2018.01.007.

[60] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine learning*, 1995, pp. 273–297, DOI: 10.1007/bf00994018.

[61] Emiliano De Cristofaro. “A Critical Overview of Privacy in Machine Learning”. In: *IEEE Security & Privacy*, 2021, pp. 19–27, DOI: 10.1109/msec.2021.3076443.

[62] James E Cutting and Lynn T Kozlowski. “Recognizing friends by their walk: Gait perception without familiarity cues”. In: *Bulletin of the psychonomic society*, 1977, pp. 353–356, DOI: 10.3758/BF03337021.

[63] Antitza Dantcheva, Petros Elia, and Arun Ross. “What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics”. In: *IEEE TIFS*, 2016, pp. 441–467, DOI: 10.1109/tifs.2015.2480381.

[64] Kourosh Darvish, Luigi Penco, Joao Ramos, Rafael Cisneros, Jerry Pratt, Eiichi Yoshida, Serena Ivaldi, and Daniele Pucci. “Teleoperation of Humanoid Robots: A Survey”. In: *IEEE Transactions on Robotics*, 2023, pp. 1706–1727, DOI: 10.1109/TRO.2023.3236952.

[65] Brendan David-John, Kevin Butler, and Eakta Jain. “For Your Eyes Only: Privacy-preserving eye-tracking datasets”. In: *Symposium on Eye Tracking Research and Applications*, 2022, pp. 1–6, DOI: 10.1145/3517031.3529618.

[66] Brendan David-John, Kevin Butler, and Eakta Jain. “Privacy-preserving datasets of eye-tracking samples with applications in XR”. In: *IEEE Transactions on Visualization and Computer Graphics*, 2023, pp. 2774–2784, DOI: 10.1109/TVCG.2023.3247048.

[67] Brendan David-John, Diane Hosfelt, Kevin Butler, and Eakta Jain. “A privacy-preserving approach to streaming eye-tracking data”. In: *IEEE Transactions on Visualization and Computer Graphics*, 2021, pp. 2555–2565, DOI: 10.1109/tvcg.2021.3067787.

[68] Essam Debie, Nour Moustafa, and Monica T. Whitty. “A Privacy-Preserving Generative Adversarial Network Method for Securing EEG Brain Signals”. In: *International Joint Conference on Neural Networks*, 2020, pp. 1–8, DOI: 10.1109/IJCNN48605.2020.9206683.

[69] Noëlie Debs, Théo Jourdan, Ali Moukadem, Antoine Boutet, and Carole Frindel. “Motion sensor data anonymization by time-frequency filtering”. In: *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1707–1711, DOI: 10.23919/Eusipco47968.2020.9287683.

[70] Najim Dehak, Patrick J Kenny, Réda Dehak, Pierre Dumouchel, and Pierre Ouellet. “Front-End Factor Analysis for Speaker Verification”. In: *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, pp. 788–798, DOI: 10.1109/tasl.2010.2064307.

[71] Silvia Del Din, Alan Godfrey, and Lynn Rochester. “Validation of an Accelerometer to Quantify a Comprehensive Battery of Gait Characteristics in Healthy Older Adults and Parkinson’s Disease: Toward Clinical and at Home Use”. In: *IEEE Journal of Biomedical and Health Informatics*, 2016, pp. 838–847, DOI: 10.1109/JBHI.2015.2419317.

[72] Jiangyi Deng, Fei Teng, Yanjiao Chen, Xiaofu Chen, Zhaohui Wang, and Wenyuan Xu. “V-Cloak: Intelligibility-, Naturalness- & Timbre-Preserving Real-Time Voice Anonymization”. In: *32nd USENIX Security Symposium, USENIX Security 2023, Anaheim, CA, USA, August 9-11, 2023*, 2023, pp. 5181–5198, URL: https://www.usenix.org/conference/usenixsecurity23/presentation/deng-jiangyi-v-cloak.

[73] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. “ArcFace: Additive Angular Margin Loss for Deep Face Recognition”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4685–4694, DOI: 10.1109/cvpr.2019.00482.

[74] Joy Derwenskus, Janet C Rucker, Alessandro Serra, John S Stahl, Deborah L Downey, Nancy L Adams, and R John Leigh. “Abnormal Eye Movements Predict Disability in MS: Two-Year Follow-Up”. In: *Annals of the New York Academy of Sciences*, 2005, pp. 521–523, DOI: 10.1196/annals.1325.058.

[75] Ding Ding, Zheyu Cao, Zhantao Gu, Hao Chen, Chang Qi, and Fang Dong. “FoANet: Focus of Attention Prediction for Foveated Pre-rendering to Enable High-quality Edge VR”. In: *ACM Transactions on Sensor Networks*, 2025, DOI: 10.1145/3722222.

[76] Apiwat Ditthapron, Emmanuel O. Agu, and Adam C. Lammert. “Privacy-Preserving Deep Speaker Separation for Smartphone-Based Passive Speech Assessment”. In: *IEEE Open J. Eng. Med. Biol.*, 2021, pp. 304–313, DOI: 10.1109/OJEMB.2021.3063994.

[77] Hamza Djelouat, Xiaojun Zhai, Mohamed Al Disi, Abbes Amira, and Faycal Bensaali. “System-on-Chip Solution for Patients Biometric: A Compressive Sensing-Based Approach”. In: *IEEE Sensors Journal*, 2018, pp. 9629–9639, DOI: 10.1109/jsen.2018.2871411.

[78] Isha Dua, Thrupthi Ann John, Riya Gupta, and CV Jawahar. “DGAZE: Driver Gaze Mapping on Road”. In: *Conference on Intelligent Robots and Systems*, 2020.

[79] Andrew T. Duchowski. *Eye Tracking Methodology*. 2017, DOI: 10.1007/978-3-319-57883-5.

[80] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. “Calibrating Noise to Sensitivity in Private Data Analysis”. In: *Journal of Privacy and Confidentiality*, 2017, pp. 17–51, DOI: 10.29012/jpc.v7i3.405.

[81] Cynthia Dwork and Aaron Roth. “The Algorithmic Foundations of Differential Privacy”. In: *Foundations and Trends® in Theoretical Computer Science*, 2013, pp. 211–407, DOI: 10.1561/0400000042.

[82] Simon Eberz, Giulio Lovisotto, Andrea Patane, Marta Kwiatkowska, Vincent Lenders, and Ivan Martinovic. “When Your Fitness Tracker Betrays You: Quantifying the Predictability of Biometric Features Across Contexts”. In: *Symposium on Security and Privacy*, 2018, pp. 889–905, DOI: 10.1109/sp.2018.00053.

[83] P. Ekman and Friesen W.V. “Facial Action Coding System”. In: *Environmental Psychology & Nonverbal Behavior*, 1978, DOI: 10.1037/t27734-000.

[84] Fatih Ertam. “An effective gender recognition approach using voice data via deeper LSTM networks”. In: *Applied Acoustics*, 2019, pp. 351–358.

[85] Bjoern M. Eskofier, Peter Federolf, Patrick F. Kugler, and Benno M. Nigg. “Marker-based classification of young–elderly gait pattern differences via direct PCA feature extraction and SVMs”. In: *Computer Methods in Biomechanics and Biomedical Engineering*, 2013, pp. 435–442, DOI: 10.1080/10255842.2011.624515.

[86] Ulrich Ettinger, Veena Kumari, Xavier A. Chitnis, Philip J. Corr, Trevor J. Crawford, Dominic G. Fannon, Séamus O’Ceallaigh, Alex L. Sumich, Victor C. Doku, and Tonmoy Sharma. “Volumetric Neural Correlates of Antisaccade Eye Movements in First-Episode Psychosis”. In: *American Journal of Psychiatry*, 2004, pp. 1918–1921, DOI: 10.1176/ajp.161.10.1918.

[87] Jiahao Fan and Xiaogang Hu. “Privacy-Preserving Motor Intent Classification via Feature Disentanglement”. In: *2023 11th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2023, pp. 1–4, DOI: 10.1109/NER52421.2023.10123842.

[88] Liyue Fan. “Image Pixelization with Differential Privacy”. In: *Data and Applications Security and Privacy XXXII*, 2018, pp. 148–162, DOI: 10.1007/978-3-319-95729-6_10.

[89] Fuming Fang, Xin Wang, Junichi Yamagishi, Isao Echizen, Massimiliano Todisco, Nicholas Evans, and Jean-Francois Bonastre. “Speaker Anonymization Using X-vector and Neural Waveform Models”. In: *Speech Synthesis Workshop*, 2019, DOI: 10.21437/ssw.2019-28.

[90] Marcos Faundez-Zanuy, Enric Sesa-Nogueras, and Stefano Marinozzi. “Speaker identification experiments under gender De-identification”. In: *Carnahan Conference on Security Technology*, 2015, pp. 1–6, DOI: 10.1109/ccst.2015.7389702.

[91] Tom Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters*, 2006, pp. 861–874, DOI: 10.1016/j.patrec.2005.10.010.

[92] Lucas Silva Figueiredo, Benjamin Livshits, David Molnar, and Margus Veanes. “Prepose: Privacy, Security, and Reliability for Gesture-Based Programming”. In: *IEEE Symposium on Security and Privacy*, 2016, pp. 122–137, DOI: 10.1109/SP.2016.16.

[93] Celia Foster, Mintao Zhao, Javier Romero, Michael J Black, Betty J Mohler, Andreas Bartels, and Isabelle Bülthoff. “Decoding subcategories of human bodies from both body-and face-responsive cortical regions”. In: *NeuroImage*, 2019, p. 116085, DOI: 10.1016/j.neuroimage.2019.116085.

[94] Lee Friedman, Mark S. Nixon, and Oleg V. Komogortsev. “Method to assess the temporal persistence of potential biometric features: Application to oculomotor, gait, face and brain structure databases”. In: *PLOS ONE*, 2017, e0178501, DOI: 10.1371/journal.pone.0178501.

[95] Wolfgang Fuhl, Efe Bozkir, and Enkelejda Kasneci. “Reinforcement learning for the privacy preservation and manipulation of eye tracking data”. In: *International Conference on Artificial Neural Networks*, 2021, pp. 595–607.

[96] Bence Galai and Csaba Benedek. “Feature selection for Lidar-based gait recognition”. In: *Workshop on Computational Intelligence for Multimedia Understanding*, 2015, pp. 1–5, DOI: 10.1109/iwcim.2015.7347076.

[97] Marco Gandolfo and Paul E Downing. “Asymmetric visual representation of sex from human body shape”. In: *Cognition*, 2020, p. 104436, DOI: 10.1016/j.cognition.2020.104436.

[98] Ana García-Blanco, Ladislao Salmerón, Manuel Perea, and Lorenzo Livianos. “Attentional biases toward emotional images in the different episodes of bipolar disorder: An eye-tracking study”. In: *Psychiatry Research*, 2014, pp. 628–633, DOI: 10.1016/j.psychres.2013.12.039.

[99] Giuseppe Garofalo, Tim Van hamme, Davy Preuveneers, and Wouter Joosen. “A Siamese Adversarial Anonymizer for Data Minimization in Biometric Applications”. In: *European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2020, pp. 334–343, DOI: 10.1109/EuroSPW51379.2020.00052.

[100] Ünal Ege Gaznepoglu and Nils Peters. “Deep Learning-based F0 Synthesis for Speaker Anonymization”. In: *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 291–295, DOI: 10.23919/EUSIPCO58844.2023.10290038.

[101] Adam Geitgey. *Face Recognition*. https://github.com/ageitgey/face_recognition, Accessed: 2021.

[102] Oana Goga, Patrick Loiseau, Robin Sommer, Renata Teixeira, and Krishna P. Gummadi. “On the Reliability of Profile Matching Across Large Online Social Networks”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, DOI: 10.1145/2783258.2788601.

[103] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative adversarial networks”. In: *Communications of the ACM*, 2020, pp. 139–144, DOI: 10.1145/3422622.

[104] Krzysztof J Gorgolewski, Tibor Auer, Vince D Calhoun, R Cameron Craddock, Samir Das, Eugene P Duff, Guillaume Flandin, Satrajit S Ghosh, Tristan Glatard, Yaroslav O Halchenko, et al. “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments”. In: *Scientific data*, 2016, pp. 1–9, DOI: 10.1038/sdata.2016.44.

[105] Yuuki Goubaru, Yasushi Yamazaki, Takeru Miyazaki, and Tetsushi Ohki. “A consideration on a common template-based biometric cryptosystem using on-line signatures”. In: *IEEE Conference on Consumer Electronics*, 2014.

[106] Eric Granger and Dmitry Gorodnichy. *Evaluation methodology for face recognition technology in video surveillance applications*. 2014.

[107] Erin Griffiths, Salah Assana, and Kamin Whitehouse. “Privacy-preserving Image Processing with Binocular Thermal Cameras”. In: *Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, pp. 1–25, DOI: 10.1145/3161198.

[108] Qiong Gui, Maria V. Ruiz-Blondet, Sarah Laszlo, and Zhanpeng Jin. “A Survey on Brain Biometrics”. In: *ACM Computing Surveys*, 2019, pp. 1–38, DOI: 10.1145/3230632.

[109] Priyanka Gupta, Gauri P. Prajapati, Shrishti Singh, Madhu R. Kamble, and Hemant A. Patil. “Design of Voice Privacy System using Linear Prediction”. In: *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2020, Auckland, New Zealand, December 7-10, 2020*, 2020, pp. 543–549, URL: https://ieeexplore.ieee.org/document/9306379.

[110] Rain Eric Haamer, Kaustubh Kulkarni, Nasrin Imanpour, Mohammad A. Haque, Egils Avots, Michelle Breisch, Kamal Nasrollahi, Sergio Escalera, Cagri Ozcinar, Xavier Baro, Ahmad R. Naghsh-Nilchi, Thomas B. Moeslund, and Golamreza Anbarjafari. “Changes in Facial Expression as Biometric: A Database and Benchmarks of Identification”. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 621–628, DOI: 10.1109/fg.2018.00098.

[111] Lindsay F Haas. “Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography”. In: *J. Neurol. Neurosurg. Psychiatry*, 2003, pp. 9–9, DOI: 10.1136/jnnp.74.1.9.

[112] Agrya Halder, Pratik Chattopadhyay, and Sathish Kumar. “Gait transformation network for gait de-identification with pose preservation”. In: *Signal, Image and Video Processing*, 2023, pp. 1753–1761.

[113] Jihun Hamm. “Enhancing utility and privacy with noisy minimax filters”. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6389–6393, DOI: 10.1109/ICASSP.2017.7953386.

[114] Yaowei Han, Sheng Li, Yang Cao, Qiang Ma, and Masatoshi Yoshikawa. “Voice-Indistinguishability: Protecting Voiceprint In Privacy-Preserving Speech Data Release”. In: *Conference on Multimedia and Expo (ICME)*, 2020, pp. 1–6, DOI: 10.1109/ICME46284.2020.9102875.

[115] Simon Hanisch, Evelyn Muschter, Admantini Hatzipanayioti, Shu-Chen Li, and Thorsten Strufe. “Understanding Person Identification Through Gait”. In: *Proceedings on Privacy Enhancing Technologies*, 2023, pp. 177–189, DOI: 10.56553/popets-2023-0011.

[116] Simon Hanisch, Loreen Pogrzeba, Evelyn Muschter, Shu-Chen Li, and Thorsten Strufe. “A kinematic dataset of locomotion with gait and sit-to-stand movements of young adults”. In: *Scientific Data*, 2024, DOI: 10.1038/s41597-024-04020-6.

[117] Hanxiang Hao, David Guera, Janos Horvath, Amy R. Reibman, and Edward J. Delp. “Robustness Analysis of Face Obscuration”. In: *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 2020, pp. 176–183, DOI: 10.1109/FG47880.2020.00021.

[118] Katarzyna Harezlak and Pawel Kasprowski. “Application of eye tracking in medicine: A survey, research issues and challenges”. In: *Computerized Medical Imaging and Graphics*, 2018, pp. 176–190, DOI: 10.1016/j.compmedimag.2017.04.006.

[119] Charles R. Harris, K. Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. “Array programming with NumPy”. In: *Nature*, 2020, pp. 357–362, DOI: 10.1038/s41586-020-2649-2.

[120] Kei Hashimoto, Junichi Yamagishi, and Isao Echizen. “Privacy-preserving sound to degrade automatic speaker verification performance”. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 5500–5504, DOI: 10.1109/ICASSP.2016.7472729.

[121] Siti Nurani Hassan, Rosnah Yusuff, Raemy Md Zein, Rizal Hussain, and Hari Krishnan Tamil Selvan. “Anthropometric data of Malaysian workers”. In: *New Ergonomics Perspective - Selected Papers of the 10th Pan-Pacific Conference on Ergonomics*, 2015, pp. 353–360, DOI: 10.1201/b17990-61.

[122] John D Herrington, Charlotte Nymberg, and Robert T Schultz. “Biological motion task performance predicts superior temporal sulcus activity”. In: *Brain and cognition*, 2011, pp. 372–381, DOI: 10.1016/j.bandc.2011.09.001.

[123] Eckhard H Hess and James M Polt. “Pupil Size as Related to Interest Value of Visual Stimuli”. In: *Science*, 1960, pp. 349–350, DOI: 10.1126/science.132.3423.349.

[124] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. “beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework”. In: *International Conference on Learning Representations*, 2017, URL: https://openreview.net/forum?id=Sy2fzU9gl.

[125] Jan Hintz, Sebastian Bayerl, Yamini Sinha, Suhita Ghosh, Martha Schubert, Sebastian Stober, Korbinian Riedhammer, and Ingo Siegert. “Anonymization of Stuttered Speech – Removing Speaker Information while Preserving the Utterance”. In: *3rd Symposium on Security and Privacy in Speech Communication*, 2023, pp. 41–45, DOI: 10.21437/SPSC.2023-7.

[126] HIPAA Compliance Assistance. *Summary of the hipaa privacy rule*. Office for Civil Rights, Accessed: 17.05.2021.

[127] Yuki Hirose, Kazuaki Nakamura, Naoko Nitta, and Noboru Babaguchi. “Anonymization of Gait Silhouette Video by Perturbing Its Phase and Shape Components”. In: *IEEE Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2019, pp. 1679–1685, DOI: 10.1109/APSIPAASC47483.2019.9023196.

[128] Thang Hoang, Deokjai Choi, and Thuc Nguyen. “Gait authentication on mobile phone using biometric cryptosystem and fuzzy commitment scheme”. In: *International Journal of Information Security*, 2015, pp. 549–560, DOI: 10.1007/s10207-015-0273-1.

[129] Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens. “An efficient P300-based brain–computer interface for disabled subjects”. In: *Journal of Neuroscience Methods*, 2008, pp. 115–125, DOI: 10.1016/j.jneumeth.2007.03.005.

[130] Philip S Holzman, Leonard R Proctor, and Dominic W Hughes. “Eye-Tracking Patterns in Schizophrenia”. In: *Science*, 1973, pp. 179–181, DOI: 10.1126/science.181.4095.179.

[131] Pei-Lun Hong, Jyun-Ya Hsiao, Chi-Hsun Chung, Yao-Min Feng, and Shun-Chi Wu. “ECG Biometric Recognition: Template-Free Approaches Based on Deep Learning”. In: *IEEE Engineering in Medicine and Biology Society*, 2019, pp. 2633–2636, DOI: 10.1109/embc.2019.8856916.

[132] Meta Horizon. *Face Tracking for Movement SDK for Unity*. https://developers.meta.com/horizon/documentation/unity/move-face-tracking/, Accessed: 2025-05-01.

[133] Fabian Horst, Sebastian Lapuschkin, Wojciech Samek, Klaus-Robert Müller, and Wolfgang I Schöllhorn. "A public dataset of overground walking kinetics and full-body kinematics in healthy individuals". In: *Mendeley Data Repository*, 2018, DOI: 10.17632/svx74xcrjr.1.

[134] Fabian Horst, Sebastian Lapuschkin, Wojciech Samek, Klaus-Robert Müller, and Wolfgang I Schöllhorn. "Explaining the unique nature of individual gait patterns with deep learning". In: *Scientific Reports*, 2019, pp. 1–13, DOI: 10.1038/s41598-019-38748-8.

[135] Syed Monowar Hossain, Amin Ahsan Ali, Md. Mahbubur Rahman, Emre Ertine David Epstein, Ashley Kennedy, Kenzie Preston, Annie Umbricht, Yixin Chen, and Santosh Kumar. "Identifying drug (cocaine) intake events from acute physiological response in the presence of free-living physical activity". In: *International Symposium on Information Processing in Sensor Networks*, 2014, pp. 71–82, DOI: 10.1109/ipsn.2014.6846742.

[136] Miao Hu, Zhenxiao Luo, Yipeng Zhou, Xuezheng Liu, and Di Wu. "Otus: A Gaze Model-based Privacy Control Framework for Eye Tracking Applications". In: *Conference on Computer Communications INFOCOM*, 2022, pp. 560–569, DOI: 10.1109/INFOCOM48880.2022.9796665.

[137] Zhiming Hu, Andreas Bulling, Sheng Li, and Guoping Wang. "EHTask: Recognizing User Tasks From Eye and Head Movements in Immersive Virtual Reality". In: *IEEE Transactions on Visualization and Computer Graphics*, 2022, pp. 1992–2004, DOI: 10.1109/tvcg.2021.3138902.

[138] Zhiming Hu, Sheng Li, Congyi Zhang, Kangrui Yi, Guoping Wang, and Dinesh Manocha. "DGaze: CNN-Based Gaze Prediction in Dynamic Scenes". In: *IEEE Transactions on Visualization and Computer Graphics*, 2020, pp. 1902–1911, DOI: 10.1109/tvcg.2020.2973473.

[139] Pei Huang, Linke Guo, Ming Li, and Yuguang Fang. "Practical Privacy-Preserving ECG-Based Authentication for IoT-Based Healthcare". In: *IEEE Internet of Things Journal*, 2019, pp. 9200–9210, DOI: 10.1109/jiot.2019.2929087.

[140] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. "DeepPrivacy: A Generative Adversarial Network for Face Anonymization". In: *International Symposium on Visual Computing*, 2019, pp. 565–578, DOI: 10.1007/978-3-030-33720-9_44.

[141] Anco Hundepool, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Eric S. Nordholt, Keith Spicer, and Peter-Paul de Wolf. *Statistical Disclosure Control*. 2012, DOI: 10.1002/9781118348239.

[142] J Thomas Hutton, JA Nagel, and Ruth B Loewenson. "Eye tracking dysfunction in Alzheimer-type dementia". In: *Neurology*, 1984, pp. 99–99, DOI: 10.1212/wnl.34.1.99.

[143] International Organization for Standardization. *Information technology — Vocabulary — Part 37: Biometrics*. Vocabulary. Geneva, CH: International Organization for Standardization, Feb. 2022.

[144] Marina Ivasic-Kos, Alexandros Iosifidis, Anastasios Tefas, and Ioannis Pitas. "Person de-identification in activity videos". In: *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, pp. 1294–1299, DOI: 10.1109/mipro.2014.6859767.

[145] M Jacquelin Perry. “Gait analysis: normal and pathological function”. In: *New Jersey: SLACK*, 2010, DOI: 10.1201/9781003525592.

[146] Salar Jafarlou, Amir M. Rahmani, Nikil Dutt, and Sanaz Rahimi Mousavi. “ECG Biosignal Deidentification Using Conditional Generative Adversarial Networks”. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2022, pp. 1366–1370, DOI: 10.1109/EMBC48229.2022.9872015.

[147] Jinhyeok Jang, Dohyung Kim, Cheonshu Park, Minsu Jang, Jaeyeon Lee, and Jaehong Kim. “ETRI-Activity3D: A Large-Scale RGB-D Dataset for Robots to Recognize Daily Activities of the Elderly”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10990–10997, DOI: 10.1109/IROS45743.2020.9341160.

[148] J Jankovic. “Parkinson’s disease: clinical features and diagnosis”. In: *J. neurol. neurosurg. psychiatry*, 2008, pp. 368–376, DOI: 10.1136/jnnp.2007.131045.

[149] Sein Jeung, Helena Cockx, Stefan Appelhoff, Timotheus Berg, Klaus Gramann, Sören Grothkopp, Elke Warmerdam, Clint Hansen, Robert Oostenveld, BIDS Maintainers, Stefan Appelhoff, Christopher J. Markiewicz, Taylor Salo, Rémi Gau, Ross Blair, Anthony Galassi, Eric Earl, Christine Rogers, Nell Hardcastle, Kimberly Ray, and Julius Welzel. “Motion-BIDS: An Extension to the Brain Imaging Data Structure to Organize Motion Data for Reproducible Research”. In: *Scientific Data*, 2024, p. 716, DOI: 10.1038/s41597-024-03559-8.

[150] Qin Jin, Arthur R. Toth, Tanja Schultz, and Alan W. Black. “Voice convergin: Speaker de-identification by voice transformation”. In: *ICASSP*, 2009, pp. 3909–3912, DOI: 10.1109/icassp.2009.4960482.

[151] Gunnar Johansson. “Visual perception of biological motion and a model for its analysis”. In: *Perception & psychophysics*, 1973, pp. 201–211, DOI: 10.3758/BF03212378.

[152] Brendan John, Ao Liu, Lirong Xia, Sanjeev Koppal, and Eakta Jain. “Let It Snow: Adding pixel noise to protect the user’s identity”. In: *Symposium on Eye Tracking Research and Applications*, 2020, pp. 1–3, DOI: 10.1145/3379157.3390512.

[153] Kerri L Johnson, Lawrie S McKay, and Frank E Pollick. “He throws like a girl (but only when he’s sad): Emotion affects sex-decoding of biological motion displays”. In: *Cognition*, 2011, pp. 265–280, DOI: 10.1016/j.cognition.2011.01.016.

[154] Théo Jourdan, Antoine Boutet, and Carole Frindel. “Toward privacy in IoT mobile devices for activity recognition”. In: *EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2018.

[155] Tadej Justin, Vitomir Struc, Simon Dobrisek, Bostjan Vesnicer, Ivo Ipsic, and France Mihelic. “Speaker de-identification using diphone recognition and speech synthesis”. In: *Automatic Face and Gesture Recognition*, 2015, pp. 1–7, DOI: 10.1109/fg.2015.7285021.

[156] E. Grace Mary Kanaga, R. Muthu Kumaran, M. Hema, R. Gowri Manohari, and Tina Anu Thomas. “An experimental investigations on classifiers for Brain Computer Interface (BCI) based authentication”. In: *Conference on Trends in Electronics and Informatics (ICEI)*, 2017, pp. 1–6, DOI: 10.1109/icoei.2017.8300873.

[157] Nader Karamzadeh, Yasaman Ardeshirpour, Matthew Kellman, Fatima Chowdhry, Afrouz Anderson, David Chorlian, Edward Wegman, and Amir Gandjbakhche. “Relative brain signature: a population-based feature extraction procedure to identify functional biomarkers in the brain of alcoholics”. In: *Brain and Behavior*, 2015, e00335, DOI: 10.1002/brb3.335.

[158] Christina Katsini, Yasmeen Abdrabou, George E. Raptis, Mohamed Khamis, and Florian Alt. “The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions”. In: *Conference on Human Factors in Computing Systems CHI*, 2020, pp. 1–21, DOI: 10.1145/3313831.3376840.

[159] Georgios Kellaris, Stavros Papadopoulos, Xiaokui Xiao, and Dimitris Papadias. “Differentially private event sequences over infinite streams”. In: *Proceedings of the VLDB Endowment*, 2014, pp. 1155–1166.

[160] Gokce Keskin, Tyler Lee, Cory Stephenson, and Oguz H. Elibol. “Measuring the Effectiveness of Voice Conversion on Speaker Identification and Automatic Speech Recognition Systems”. In: *arXiv:1905.12531 [eess]*, 2019, DOI: 10.48550/arxiv.1905.12531.

[161] W Khalifa, A Salem, and M Roushdy. “A Survey of EEG Based User Authentication Schemes”. In: *International Conference on INFOrmatics and Systems*, 2012, pp. 55–60.

[162] Siddhartha Khandelwal and Nicholas Wickström. “Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database”. In: *Gait & posture*, 2017, pp. 84–90, DOI: 10.1016/j.gaitpost.2016.09.023.

[163] Łukasz Kidziński, Bryan Yang, Jennifer L Hicks, Apoorva Rajagopal, Scott L Delp, and Michael H Schwartz. “Deep neural networks enable quantitative movement analysis using single-camera videos”. In: *Nature communications*, 2020, pp. 1–10, DOI: 10.1038/s41467-020-17807-z.

[164] Diederik P Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: 2022, DOI: 10.48550/arxiv.1312.6114.

[165] Christopher Kirtley. *Clinical gait analysis: theory and practice*. 2006, ISBN: 978-0-443-10009-3.

[166] Barbara Kitchenham. “Procedures for Performing Systematic Reviews”. In: *Keele, UK, Keele Univ.*, 2004.

[167] Kazuhiro Kondo, Tomohiro Komiyama, and Shintaro Kashiwada. “Towards Gender-Dependent Babble Maskers for Speech Privacy Protection”. In: *IEEE Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2013, pp. 275–278, DOI: 10.1109/iih-msp.2013.77.

[168] Kazuhiro Kondo and Hiroki Sakurai. “Gender-Dependent Babble Maskers Created from Multi-speaker Speech for Speech Privacy Protection”. In: *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2014, pp. 251–254, DOI: 10.1109/iih-msp.2014.69.

[169] Thomas Kopalidis, Vassilios Solachidis, Nicholas Vretos, and Petros Daras. “Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets”. In: *Information*, 2024, p. 135, DOI: 10.3390/info15030135.

[170] Lynn T Kozlowski and James E Cutting. “Recognizing the sex of a walker from a dynamic point-light display”. In: *Perception & psychophysics*, 1977, pp. 575–580, DOI: 10.3758/BF03198740.

[171] Mark A Kramer. “Nonlinear principal component analysis using autoassociative neural networks”. In: *AIChE journal*, 1991, pp. 233–243.

[172] Krzysztof Krejtz, Andrew T. Duchowski, Anna Niedzielska, Cezary Biele, and Izabela Krejtz. “Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze”. In: *PLOS ONE*, 2018, e0203629, DOI: 10.1371/journal.pone.0203629.

[173] Jacob Leon Kröger, Otto Hans-Martin Lutz, and Florian Müller. “What Does Your Gaze Reveal About You? On the Privacy Implications of Eye Tracking”. In: *Privacy and Identity Management. Data for Better Living: AI and Privacy*, 2020, pp. 226–241, DOI: 10.1007/978-3-030-42504-3_15.

[174] Craig A Kuechenmeister, Patrick H Linton, Thelma V Mueller, and Hilton B White. “Eye Tracking in Relation to Age, Sex, and Illness”. In: *Arch. Gen. Psychiatry*, 1977, pp. 578–579, DOI: 10.1001/archpsyc.1977.01770170088008.

[175] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. In: *The Annals of Mathematical Statistics*, 1951, pp. 79–86, DOI: 10.1214/aoms/1177729694.

[176] Joachim Lange, Karsten Georg, and Markus Lappe. “Visual perception of biological motion by form: A template-matching analysis”. In: *Journal of vision*, 2006, pp. 6–6, DOI: 10.1167/6.8.6.

[177] Minh-Ha Le and Niklas Carlsson. “StyleID: Identity Disentanglement for Anonymizing Faces”. In: *Proceedings on Privacy Enhancing Technologies*, 2023, pp. 264–278, DOI: 10.56553/popets-2023-0016.

[178] Minh-Ha Le, Md Sakib Nizam Khan, Georgia Tsaloli, Niklas Carlsson, and Sonja Buchegger. “AnonFACES: Anonymizing Faces Adjusted to Constraints on Efficacy and Security”. In: *Proceedings of the 19th Workshop on Privacy in the Electronic Society*, 2020, DOI: 10.1145/3411497.3420220.

[179] Jin Pyo Lee, Hanhyeok Jang, Yeonwoo Jang, Hyeonseo Song, Suwoo Lee, Pooi See Lee, and Jiyun Kim. “Encoding of multi-modal emotional information via personalized skin-integrated wireless facial interface”. In: *Nature Communications*, 2024, DOI: 10.1038/s41467-023-44673-2.

[180] Juho Leinonen, Petri Ihantola, and Arto Hellas. “Preventing Keystroke Based Identification in Open Data Sets”. In: *ACM Conference on Learning @ Scale*, 2017.

[181] Don S Lemons and Paul Langevin. *An introduction to stochastic processes in physics*. 2002.

[182] Deborah L. Levy, Anne B. Sereno, Diane C. Gooding, and Gilllian A. O’Driscoll. “Eye Tracking Dysfunction in Schizophrenia: Characterization and Pathophysiology”. In: *Behavioral Neurobiology of Schizophrenia and Its Treatment*. Springer, 2010, pp. 311–347. DOI: 10.1007/7854_2010_60.

[183] Jingjie Li, Amrita Roy Chowdhury, Kassem Fawaz, and Younghyun Kim. “Kalϵdo: Real-Time Privacy Control for Eye-Tracking Systems”. In: *USENIX Security*, 2021, pp. 1793–1810, URL: https://www.usenix.org/conference/usenixsecurity21/presentation/li-jingjie.

[184] Jonathan Liebers, Patrick Horn, Christian Burschik, Uwe Gruenefeld, and Stefan Schneegass. "Using Gaze Behavior and Head Orientation for Implicit Identification in Virtual Reality". In: *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, 2021, pp. 1–9, DOI: 10.1145/3489849.3489880.

[185] Jae Lim and A. Oppenheim. "All-pole modeling of degraded speech". In: *Transactions on Audio, Speech, and Language Processing*, 1978, pp. 197–210, DOI: 10.1109/tassp.1978.1163086.

[186] Ana Lígia Silva de Lima, Luc J. W. Evers, Tim Hahn, Lauren Bataille, Jamie L. Hamilton, Max A. Little, Yasuyuki Okuma, Bastiaan R. Bloem, and Marjan J. Faber. "Freezing of gait and fall detection in Parkinson's disease using wearable sensors: a systematic review". In: *Journal of Neurology*, 2017, pp. 1642–1654, DOI: 10.1007/s00415-017-8424-0.

[187] Ao Liu, Lirong Xia, Andrew Duchowski, Reynold Bailey, Kenneth Holmqvist, and Eakta Jain. "Differential privacy for eye-tracking data". In: *ACM Symposium on Eye Tracking Research & Applications*, 2019, pp. 1–10, DOI: 10.1145/3314111.3319823.

[188] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C Kot. "NTU RGB+D 120: A large-scale benchmark for 3D human activity understanding". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, pp. 2684–2701.

[189] Xinwen Liu, Huan Wang, Zongjin Li, and Lang Qin. "Deep learning in ECG diagnosis: A review". In: *Knowledge-Based Systems*, 2021, p. 107187, DOI: 10.1016/j.knosys.2021.107187.

[190] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. "Deep Learning Face Attributes in the Wild". In: *Proceedings of the IEEE international conference on computer vision*, 2015, DOI: 10.1109/iccv.2015.425.

[191] Florian Loffing, Florian Sölter, Norbert Hagemann, and Bernd Strauss. "On-court position and handedness in visual anticipation of stroke direction in tennis". In: *Psychology of Sport and Exercise*, 2016, pp. 195–204, DOI: 10.1016/j.psychsport.2016.08.014.

[192] Dillon Lohr, Samantha Aziz, Lee Friedman, and Oleg V Komogortsev. "GazeBaseVR, a large-scale, longitudinal, binocular eye-tracking dataset collected in virtual reality". In: *Scientific Data*, 2023, p. 177, DOI: https://doi.org/10.1038/s41597-023-02075-5.

[193] Dillon Lohr, Henry Griffith, Samantha Aziz, and Oleg Komogortsev. "A Metric Learning Approach to Eye Movement Biometrics". In: *2020 IEEE International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–7, DOI: 10.1109/IJCB48548.2020.9304859.

[194] Dillon Lohr and Oleg V. Komogortsev. "Eye Know You Too: Toward Viable End-to-End Eye Movement Biometrics for User Authentication". In: *IEEE Transactions on Information Forensics and Security*, 2022, pp. 3151–3164, DOI: 10.1109/tifs.2022.3201369.

[195] Harald Loose and Jon Lindström Bolmgren. "GaitAnalysisDataBase–Short Overview". In: *Tech. Hochsch. Brandenbg*, 2019, pp. 1–6.

[196] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. “SMPL: a skinned multi-person linear model”. In: *ACM Transactions on Graphics*, 2015, pp. 1–16, DOI: 10.1145/2816795.2818013.

[197] Paula Lopez-Otero, Carmen Magariños, Laura Docio-Fernandez, Eduardo Rodriguez-Banga, Daniel Erro, and Carmen Garcia-Mateo. “Influence of speaker de-identification in depression detection”. In: *IET signal process.*, 2017.

[198] Elena Lorenzi, Uwe Mayer, Orsola Rosa-Salva, and Giorgio Vallortigara. “Dynamic features of animate motion activate septal and preoptic areas in visually naïve chicks (Gallus gallus)”. In: *Neuroscience*, 2017, pp. 54–68, DOI: 10.1016/j.neuroscience.2017.04.022..

[199] Viktor Losing and Martina Hasenjäger. “A multi-modal gait database of natural everyday-walk in an urban environment”. In: *Scientific Data*, 2022, p. 473, DOI: 10.1038/s41597-022-01580-3.

[200] Fani Loula, Sapna Prasad, Kent Harber, and Maggie Shiffrar. “Recognizing people from their movement”. In: *Journal of Experimental Psychology: Human Perception and Performance*, 2005, p. 210, DOI: 10.1037/0096-1523.31.1.210.

[201] Li Lu, Jiadi Yu, Yingying Chen, Hongbo Liu, Yanmin Zhu, Yunfei Liu, and Minglu Li. “LipPass: Lip Reading-based User Authentication on Smartphones Leveraging Acoustic Signals”. In: *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 1466–1474, DOI: 10.1109/INFOCOM.2018.8486283.

[202] Yue Luo, Sarah M Coppola, Philippe C Dixon, Song Li, Jack T Dennerlein, and Boyi Hu. “A database of human gait performance on irregular and uneven surfaces collected by wearable sensors”. In: *Scientific data*, 2020, p. 219, DOI: 10.1038/s41597-020-0563-y.

[203] Yuanjun Lv, Jixun Yao, Peikun Chen, Hongbin Zhou, Heng Lu, and Lei Xie. “Salt: Distinguishable Speaker Anonymization Through Latent Space Transformation”. In: *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2023, pp. 1–8.

[204] Xiaosong Ma, Yubo Song, Zhongwei Wang, Shang Gao, Bin Xiao, and Aiqun Hu. “You Can Hear But You Cannot Record: Privacy Protection by Jamming Audio Recording”. In: *International Conference on Communications*, 2021, pp. 1–6, DOI: 10.1109/ICC42927.2021.9500456.

[205] Carmen Magariños, Paula Lopez-Otero, Laura Docio-Fernandez, Eduardo Rodriguez-Banga, Daniel Erro, and Carmen Garcia-Mateo. “Reversible speaker de-identification using pre-trained transformation functions”. In: *Computer Speech & Language*, 2017, pp. 36–52, DOI: 10.1016/j.csl.2017.05.001.

[206] Ahmed Mahfouz, Tarek M. Mahmoud, and Ahmed Sharaf Eldin. “A survey on behavioral biometric authentication on smartphones”. In: *Journal of Information Security and Applications*, 2017, pp. 28–37, DOI: 10.1016/j.jisa.2017.10.002.

[207] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. “AMASS: Archive of Motion Capture As Surface Shapes”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 5442–5451, DOI: 10.1109/iccv.2019.00554.

[208] Seedahmed S. Mahmoud. “A generalised wavelet packet-based anonymisation approach for ECG security application”. In: *Security and Communication Networks*, 2016, pp. 6137–6147, DOI: 10.1002/sec.1762.

[209] Emanuele Maiorana, Patrizio Campisi, and Alessandro Neri. “Bioconvolving: Cancelable templates for a multi-biometrics signature recognition system”. In: *IEEE International Systems Conference*, 2011.

[210] Anindya Maiti, Oscar Armbruster, Murtuza Jadliwala, and Jibo He. “Smartwatch-Based Keystroke Inference Attacks and Context-Aware Protection Mechanisms”. In: *ACM Asia Conference on Computer and Communications Security*, 2016.

[211] Paivi Majaranta and Andreas Bulling. “Eye Tracking and Eye-Based Human–Computer Interaction”. In: *Human–Computer Interaction*. Springer London, 2014, pp. 39–65. DOI: 10.1007/978-1-4471-6392-3_3.

[212] Matthew Malek–Podjaski and Fani Deligianni. “Towards Explainable, Privacy-Preserved Human-Motion Affect Recognition”. In: *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2021, pp. 01–09, DOI: 10.1109/SSCI50451.2021.9660129.

[213] Mohammad Malekzadeh, Richard G. Clegg, Andrea Cavallaro, and Hamed Haddadi. “Privacy and utility preserving sensor-data transformations”. In: *Pervasive and Mobile Computing*, 2020, p. 101132, DOI: 10.1016/j.pmcj.2020.101132.

[214] Valeria Manera, Ben Schouten, Cristina Becchio, Bruno G Bara, and Karl Verfaillie. “Inferring intentions from biological motion: a stimulus set of point-light communicative interactions”. In: *Behavior research methods*, 2010, pp. 168–178, DOI: 10.3758/BRM.42.1.168.

[215] M. Sabarimalai Manikandan and S. Dandapat. “ECG Distortion Measures and their Effectiveness”. In: *Emerging Trends in Engineering and Technology*, 2008, pp. 705–710, DOI: 10.1109/icetet.2008.248.

[216] I. Martinovic, D. Davies, M. Frank, D. Perito, T. Ros, and D. Song. “On the Feasibility of Side-Channel Attacks with Brain-Computer Interfaces”. In: *USENIX Security*, 2012, pp. 143–158.

[217] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. “DeepLabCut: markerless pose estimation of user-defined body parts with deep learning”. In: *Nature neuroscience*, 2018, pp. 1281–1289, DOI: 10.1038/s41593-018-0209-y.

[218] Richard Matovu and Abdul Serwadda. “Your substance abuse disorder is an open secret! Gleaning sensitive personal information from templates in an EEG-based authentication system”. In: *IEEE Conference on Biometrics Theory, Applications and Systems*, 2016, pp. 1–7, DOI: 10.1109/btas.2016.7791210.

[219] Richard Matovu, Abdul Serwadda, David Irakiza, and Isaac Griswold-Steiner. “Jekyll and Hyde: On The Double-Faced Nature of Smart-Phone Sensor Noise Injection”. In: *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2018, pp. 1–6, DOI: 10.23919/BIOSIG.2018.8553043.

[220] Gerald Matthews, W Middleton, Bernard Gilmartin, and Mark A Bullimore. “Pupillary diameter and cognitive load.” In: *Journal of Psychophysiology*, 1991, pp. 265–271.

[221] Candy Olivia Mawalim, Kasorn Galajit, Jessada Karnjana, Shunsuke Kidani, and Masashi Unoki. “Speaker anonymization by modifying fundamental frequency and x-vector singular value”. In: *Computer Speech & Language*, 2022, p. 101326, DOI: 10.1016/j.csl.2021.101326.

[222] Candy Olivia Mawalim, Shogo Okada, and Masashi Unoki. “Speaker anonymization by pitch shifting based on time-scale modification”. In: *Proc. 2nd Symp. Secur. Privacy Speech Commun*, 2022, pp. 35–42.

[223] Maxim Maximov, Ismail Elezi, and Laura Leal-Taixé. “CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks”. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5446–5455, DOI: 10.1109/CVPR42600.2020.00549.

[224] Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. “Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi”. In: *Proc. Interspeech 2017*, 2017, pp. 498–502, DOI: 10.21437/Interspeech.2017-1386.

[225] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. “Defeating Image Obfuscation with Deep Learning”. In: *arXiv:1609.00408 [cs]*, 2016, URL: http://arxiv.org/abs/1609.00408.

[226] Frank D. McSherry. “Privacy Integrated Queries: An Extensible Platform for Privacy-preserving Data Analysis”. In: *SIGMOD*, 2009, pp. 19–30, DOI: 10.1145/1559845.1559850.

[227] Blaž Meden, Peter Rot, Philipp Terhörst, Naser Damer, Arjan Kuijper, Walter J. Scheirer, Arun Ross, Peter Peer, and Vitomir Štruc. “Privacy–enhancing face biometrics: A comprehensive survey”. In: *IEEE Transactions on Information Forensics and Security*, 2021, pp. 4147–4183, DOI: 10.1109/TIFS.2021.3096024.

[228] Sina Mehdizadeh, Hoda Nabavi, Andrea Sabo, Twinkle Arora, Andrea Iaboni, and Babak Taati. “Concurrent validity of human pose tracking in video for measuring gait parameters in older adults: a preliminary analysis with multiple trackers, viewing angles, and walking directions”. In: *Journal of NeuroEngineering and Rehabilitation*, 2021, pp. 1–16, DOI: 10.1186/s12984-021-00933-0.

[229] Lubin Meng, Xue Jiang, Jian Huang, Wei Li, Hanbin Luo, and Dongrui Wu. “User Identity Protection in EEG-Based Brain–Computer Interfaces”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023, pp. 3576–3586, DOI: 10.1109/TNSRE.2023.3310883.

[230] Weizhi Meng, Duncan S. Wong, Steven Furnell, and Jianying Zhou. “Surveying the Development of Biometric User Authentication on Mobile Phones”. In: *IEEE Communications Surveys & Tutorials*, 2015, pp. 1268–1293, DOI: 10.1109/comst.2014.2386915.

[231] Yan Meng, Yuxia Zhan, Jiachun Li, Suguo Du, Haojin Zhu, and Xuemin Shen. “De-Anonymizing Avatars in Virtual Reality: Attacks and Countermeasures”. In: *IEEE Transactions on Mobile Computing*, 2024, pp. 13342–13357, DOI: 10.1109/TMC.2024.3426046.

[232] *Meta Quest Pro*. https://en.wikipedia.org/wiki/Meta_Quest_Pro, Accessed: 2025-03-03.

[233] Sarina Meyer, Florian Lux, Pavel Denisov, Julia Koch, Pascal Tilli, and Ngoc Thang Vu. “Speaker Anonymization with Phonetic Intermediate Representations”. In: *Interspeech 2022*, 2022, pp. 4925–4929, DOI: 10.21437/Interspeech.2022-10703.

[234] Sarina Meyer, Florian Lux, Julia Koch, Pavel Denisov, Pascal Tilli, and Ngoc Thang Vu. “Prosody Is Not Identity: A Speaker Anonymization Approach Using Prosody Cloning”. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5, DOI: 10.1109/ICASSP49357.2023.10096607.

[235] Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, and Natalia Tomashenko. “Language-Independent Speaker Anonymization Approach Using Self-Supervised Pre-Trained Models”. In: *The Speaker and Language Recognition Workshop (Odyssey 2022)*, 2022, pp. 279–286, DOI: 10.21437/Odyssey.2022-39.

[236] Xiaoxiao Miao, Xin Wang, Erica Cooper, Junichi Yamagishi, and Natalia Tomashenko. “Speaker Anonymization Using Orthogonal Householder Neural Network”. In: *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, 2023, pp. 3681–3695, DOI: 10.1109/TASLP.2023.3313429.

[237] Denis Migdal and Christophe Rosenberger. “Keystroke Dynamics Anonymization System”. In: *Joint Conference on e-Business and Telecommunications*, 2019.

[238] Denis Migdal and Christophe Rosenberger. “My Behavior is my Privacy & Secure Password !” In: *IEEE Conference on Cyberworlds*, 2019.

[239] Senay Mihcin. “Simultaneous validation of wearable motion capture system for lower body applications: over single plane range of motion (ROM) and gait activities”. In: *Biomedical Engineering / Biomedizinische Technik*, 2022, pp. 185–199, DOI: 10.1515/bmt-2021-0429.

[240] Mark Roman Miller, Fernanda Herrera, Hanseul Jun, James A Landay, and Jeremy N Bailenson. “Personal identifiability of user tracking data during observation of 360-degree VR video”. In: *Scientific Reports*, 2020, p. 17404, DOI: 10.1038/s41598-020-74486-y.

[241] *Mocopi*. https://electronics.sony.com/more/mocopi/all-mocopi/p/qmss1-uscx, Accessed: 2025-03-03.

[242] Joann M Montepare and Leslie Zebrowitz-McArthur. “Impressions of people created by age-related qualities of their gaits”. In: *Journal of personality and social psychology*, 1988, p. 547, DOI: 10.1037/0022-3514.55.4.547.

[243] G.B. Moody and R.G. Mark. “The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it”. In: *Proceedings Computers in Cardiology*, 1990, pp. 185–188, DOI: 10.1109/cic.1990.144205.

[244] Saemi Moon, Myeonghyeon Kim, Zhenyue Qin, Yang Liu, and Dongwoo Kim. “Anonymization for skeleton action recognition”. In: *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, 2023, pp. 15028–15036, DOI: 10.1609/aaai.v37i12.26754.

[245] Alec G. Moore, Ryan P. McMahan, Hailiang Dong, and Nicholas Ruozzi. "Personal Identifiability and Obfuscation of User Tracking Data From VR Training Sessions". In: *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2021, pp. 221–228, DOI: 10.1109/ISMAR52148.2021.00037.

[246] Goncalo Moreira, Andre Graca, Bruno Silva, Pedro Martins, and Jorge Batista. "Neuromorphic Event-based Face Identity Recognition". In: *2022 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 922–929, DOI: 10.1109/icpr56361.2022.9956236.

[247] Edward R Morrison, Hannah Bain, Louise Pattison, and Hannah Whyte-Smith. "Something in the way she moves: biological motion, body shape, and attractiveness in women". In: *Visual Cognition*, 2018, pp. 405–411, DOI: 10.1080/13506285.2018.1471560.

[248] Aymen Mtibaa, Dijana Petrovska-Delacretaz, and Ahmed Ben Hamida. "Cancelable speaker verification system based on binary Gaussian mixtures". In: *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2018, pp. 1–6, DOI: 10.1109/ATSIP.2018.8364513.

[249] Naoya Mukojima, Masaki Yasugi, Yasuhiro Mizutani, Takeshi Yasui, and Hirotsugu Yamamoto. "Deep-Learning-Assisted Single-Pixel Imaging for Gesture Recognition in Consideration of Privacy". In: *IEICE Transactions on Electronics*, 2022, pp. 79–85, DOI: 10.1587/transele.2021dii0002.

[250] Vivek Nair, Wenbo Guo, Justus Mattern, Rui Wang, James F O'Brien, Louis Rosenberg, and Dawn Song. "Unique identification of 50,000+ virtual reality users from head & hand motion data". In: *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 895–910.

[251] Vivek Nair, Wenbo Guo, James F. O'Brien, Louis Rosenberg, and Dawn Song. "Deep Motion Masking for Secure, Usable, and Scalable Real-Time Anonymization of Ecological Virtual Reality Motion Data". In: *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2024, pp. 493–500, DOI: 10.1109/VRW62533.2024.00096.

[252] Vivek Nair, Mark Roman Miller, Rui Wang, Brandon Huang, Christian Rack, Marc Erich Latoschik, and James F. O'Brien. "Effect of Data Degradation on Motion Re-Identification". In: *2024 IEEE 25th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2024, pp. 85–90, DOI: 10.1109/WoWMoM60985.2024.00026.

[253] Vivek C Nair, Gonzalo Munilla-Garrido, and Dawn Song. "Going Incognito in the Metaverse: Achieving Theoretically Optimal Privacy-Usability Tradeoffs in VR". In: *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 2023, DOI: 10.1145/3586183.3606754.

[254] Andreas Nautsch, Jose Patino, Natalia Tomashenko, Junichi Yamagishi, Paul-Gauthier Noe, Jean-Francois Bonastre, Massimiliano Todisco, and Nicholas Evans. "The Privacy ZEBRA: Zero Evidence Biometric Recognition Assessment". In: *Interspeech*, 2020, DOI: 10.21437/interspeech.2020-1815.

[255] Alexandru Nelus and Rainer Martin. "Gender Discrimination Versus Speaker Identification Through Privacy-Aware Adversarial Feature Extraction". In: *Speech Communication; 13th ITG-Symposium*, 2018, pp. 1–5.

[256] Alexandru Nelus and Rainer Martin. “Privacy-Preserving Audio Classification Using Variational Information Feature Extraction”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, pp. 2864–2877, DOI: 10.1109/taslp.2021.3108063.

[257] Francesco Nespoli, Daniel Barreda, Jöerg Bitzer, and Patrick A. Naylor. “Two-Stage Voice Anonymization for Enhanced Privacy”. In: *INTERSPEECH 2023*, 2023, pp. 3854–3858, DOI: 10.21437/Interspeech.2023-1341.

[258] SUNY Downstate Medical Center Neurodynamics Laboratory. *EEG Database*. http://kdd.ics.uci.edu/databases/eeg/eeg.data.html, Accessed: 2021.

[259] Elaine M Newton, Latanya Sweeney, and Bradley Malin. “Preserving privacy by de-identifying face images”. In: *IEEE Transactions on Knowledge and Data Engineering*, 2005, pp. 232–243, DOI: 10.1109/tkde.2005.32.

[260] Paul-Gauthier Noé, Jean-François Bonastre, Driss Matrouf, N. Tomashenko, Andreas Nautsch, and Nicholas Evans. “Speech Pseudonymisation Assessment Using Voice Similarity Matrices”. In: *Interspeech*, 2020, DOI: 10.21437/interspeech.2020-2720.

[261] Paul-Gauthier Noé, Andreas Nautsch, Nicholas Evans, Jose Patino, Jean-François Bonastre, Natalia Tomashenko, and Driss Matrouf. “Towards a unified assessment framework of speech pseudonymisation”. In: *Computer Speech & Language*, 2022, p. 101299, DOI: 10.1016/j.csl.2021.101299.

[262] Alexis Nolin-Lapalme, Robert Avram, and Hussin Julie. “PrivECG: generating private ECG for end-to-end anonymization”. In: *Machine Learning for Healthcare Conference*, 2023, pp. 509–528.

[263] Iyad Obeid and Joseph Picone. “The temple university hospital EEG data corpus”. In: *Frontiers in neuroscience*, 2016, p. 196.

[264] Ikenna Odinaka, Po-Hsiang Lai, Alan D. Kaplan, Joseph A. O’Sullivan, Erik J. Sirevaag, and John W. Rohrbaugh. “ECG Biometric Recognition: A Comparative Analysis”. In: *IEEE TIFS*, 2012, pp. 1812–1824, DOI: 10.1109/tifs.2012.2215324.

[265] Yoshitaka Ohshio, Haruka Adachi, Kenta Iwai, Takanobu Nishiura, and Yoichi Yamashita. “Active Speech Obscuration with Speaker-dependent Human Speech-like Noise for Speech Privacy”. In: *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2018, pp. 1252–1255, DOI: 10.23919/APSIPA.2018.8659754.

[266] Michele Panariello, Francesco Nespoli, Massimiliano Todisco, and Nicholas Evans. “Speaker anonymization using neural audio codec language models”. In: *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 4725–4729.

[267] Julien Pansiot, Danail Stoyanov, Douglas McIlwraith, Benny P.L. Lo, and G. Z. Yang. “Ambient and Wearable Sensor Fusion for Activity Recognition in Healthcare Monitoring Systems”. In: *Workshop on Wearable and Implantable Body Sensor Networks*, 2007, pp. 208–212, DOI: 10.1007/978-3-540-70994-7_36.

[268] Liuba Papeo, Moritz F Wurm, Nikolaas N Oosterhof, and Alfonso Caramazza. “The neural representation of human versus nonhuman bipeds and quadrupeds”. In: *Scientific reports*, 2017, pp. 1–8, DOI: 10.1038/s41598-017-14424-7.

[269] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman. “Deep Face Recognition”. In: *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*, 2015, pp. 41.1–41.12, DOI: 10.5244/c.29.41.

[270] Sree Hari Krishnan Parthasarathi, H. Bourlard, and D. Gatica-Perez. “Wordless Sounds: Robust Speaker Diarization Using Privacy-Preserving Audio Representations”. In: *IEEE Transactions on Audio, Speech, and Language Processing*, 2013, pp. 85–98, DOI: 10.1109/tasl.2012.2215588.

[271] Sree Hari Krishnan Parthasarathi, Hervé Bourlard, and Daniel Gatica-Perez. “LP residual features for robust, privacy-sensitive speaker diarization”. In: *Interspeech*, 2011.

[272] Sree Hari Krishnan Parthasarathi, Mathew Magimai.-Doss, Daniel Gatica-Perez, and Hervé Bourlard. “Speaker change detection with privacy-preserving audio cues”. In: *Proceedings of the 2009 international conference on Multimodal interfaces - ICMI-MLMI '09*, 2009, p. 343, DOI: 10.1145/1647314.1647385.

[273] Damian Pascual, Alireza Amirshahi, Amir Aminifar, David Atienza, Philippe Ryvlin, and Roger Wattenhofer. “EpilepsyGAN: Synthetic Epileptic Brain Activities With Privacy Preservation”. In: *IEEE Transactions on Biomedical Engineering*, 2021, pp. 2435–2446, DOI: 10.1109/tbme.2020.3042574.

[274] Manas A. Pathak and Bhiksha Raj. “Privacy-preserving speaker verification as password matching”. In: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1849–1852, DOI: 10.1109/icassp.2012.6288262.

[275] Jose Patino, Natalia Tomashenko, Massimiliano Todisco, Andreas Nautsch, and Nicholas Evans. “Speaker Anonymisation Using the McAdams Coefficient”. In: *Interspeech*, 2021, pp. 1099–1103, DOI: 10.21437/Interspeech.2021-1070.

[276] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. “Expressive Body Capture: 3D Hands, Face, and Body from a Single Image”. In: *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[277] Karl Pearson. “LIII. On lines and planes of closest fit to systems of points in space”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, pp. 559–572, DOI: 10.1080/14786440109462720.

[278] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research*, 2011, pp. 2825–2830, URL: http://scikit-learn.sourceforge.net.

[279] Yujia Peng, Hannah Lee, Tianmin Shu, and Hongjing Lu. “Exploring biological motion perception in two-stream convolutional neural networks”. In: *Vision Research*, 2020, pp. 28–40, DOI: 10.1016/j.visres.2020.09.005.

[280] Juan M. Perero-Codosero, Fernando M. Espinoza-Cuadros, and Luis A. Hernández-Gómez. “X-vector anonymization using autoencoders and adversarial training for preserving speech privacy”. In: *Comput. Speech Lang.*, 2022, DOI: 10.1016/j.csl.2022.101351.

[281] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian 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*, 2019, pp. 1–12, DOI: 10.1145/3290605.3300340.

[282] Esteban Piacentino and Cecilio Angulo. “Generating fake data using GANs for anonymizing healthcare data”. In: *International Work-Conference on Bioinformatics and Biomedical Engineering*, 2020, pp. 406–417.

[283] R. Plamondon and S.N. Srihari. “Online and off-line handwriting recognition: a comprehensive survey”. In: *IEEE TPAMI*, 2000, pp. 63–84, DOI: 10.1109/34.824821.

[284] Kurt Plarre, Andrew Raij, Syed Monowar Hossain, Amin Ahsan Ali, Motohiro Nakajima, Mustafa Al’absi, Emre Ertin, Thomas Kamarck, Santosh Kumar, Marcia Scott, Daniel Siewiorek, Asim Smailagic, and Lorentz E. Wittmers. “Continuous inference of psychological stress from sensory measurements collected in the natural environment”. In: *International Conference on Information Processing in Sensor Networks*, 2011, pp. 97–108.

[285] M. Pobar and I. Ipsic. “Online speaker de-identification using voice transformation”. In: *IEEE Convention on Information and Communication Technology, Electronics and Microelectronics*, 2014, pp. 1264–1267, DOI: 10.1109/mipro.2014.6859761.

[286] Bogdan Pogorelc, Zoran Bosnić, and Matjaž Gams. “Automatic recognition of gait-related health problems in the elderly using machine learning”. In: *Multimedia Tools and Applications*, 2011, pp. 333–354, DOI: 10.1007/s11042-011-0786-1.

[287] Loreen Pogrzeba, Evelyn Muschter, Simon Hanisch, Veronica Y. P. Wardhani, Thorsten Strufe, Frank H. P. Fitzek, and Shu-Chen Li. “A Full-Body IMU-Based Motion Dataset of Daily Tasks by Older and Younger Adults”. In: *Scientific Data*, 2025, p. 531, DOI: 10.1038/s41597-025-04818-y.

[288] F. Pollick, J. Kay, K. Heim, and R. Stringer. “Gender recognition from point-light walkers.” In: *J Exp Psychol Hum Percept Perform*, 2005.

[289] Alex Poole and Linden J. Ball. “Eye Tracking in HCI and Usability Research”. In: *Encyclopedia of Human Computer Interaction*. IGI Global, 2006, pp. 211–219. DOI: 10.4018/978-1-59140-562-7.ch034.

[290] Jose Portelo, Alberto Abad, Bhiksha Raj, and Isabel Trancoso. “Secure Binary Embeddings of Front-End Factor Analysis for Privacy Preserving Speaker Verification”. In: *Interspeech*, 2013, pp. 2494–2498.

[291] Jose Portelo, Bhiksha Raj, Alberto Abad, and Isabel Trancoso. “Privacy-preserving speaker verification using garbled GMMS”. In: *European Signal Processing Conference*, 2014.

[292] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. “The Kaldi speech recognition toolkit”. In: *Workshop on automatic speech recognition and understanding*, 2011.

[293] Gauri P. Prajapati, Dipesh K. Singh, Preet P. Amin, and Hemant A. Patil. “Voice Privacy Through x-Vector and CycleGAN-Based Anonymization”. In: *Interspeech*, 2021, pp. 1684–1688, DOI: 10.21437/Interspeech.2021-1573.

[294] Gauri P. Prajapati, Dipesh K. Singh, Preet P. Amin, and Hemant A. Patil. “Voice privacy using CycleGAN and time-scale modification”. In: *Comput. Speech Lang.*, 2022, DOI: 10.1016/j.csl.2022.101353.

[295] Jiří Přibil, Anna Přibilová, and Jindřich Matoušek. “Evaluation of speaker de-identification based on voice gender and age conversion”. In: *Journal of Electrical Engineering*, 2018, pp. 138–147, DOI: 10.2478/jee-2018-0017.

[296] Python Core Team. *Python: A dynamic, open source programming language*. Python version 3.8.3. Python Software Foundation. 2019. URL: https://www.python.org/.

[297] Jianwei Qian, Haohua Du, Jiahui Hou, Linlin Chen, Taeho Jung, and Xiang-Yang Li. “Hidebehind: Enjoy Voice Input with Voiceprint Unclonability and Anonymity”. In: *ACM Conference on Embedded Networked Sensor Systems*, 2018, pp. 82–94, DOI: 10.1145/3274783.3274855.

[298] Jianwei Qian, Haohua Du, Jiahui Hou, Linlin Chen, Taeho Jung, and Xiangyang Li. “Speech Sanitizer: Speech Content Desensitization and Voice Anonymization”. In: *IEEE Transactions on Dependable and Secure Computing*, 2021, pp. 2631–2642, DOI: 10.1109/tdsc.2019.2960239.

[299] Jianwei Qian, Feng Han, Jiahui Hou, Chunhong Zhang, Yu Wang, and Xiang-Yang Li. “Towards Privacy-Preserving Speech Data Publishing”. In: *INFOCOM*, 2018, pp. 1079–1087, DOI: 10.1109/infocom.2018.8486250.

[300] Yaron Rachlin and Dror Baron. “The secrecy of compressed sensing measurements”. In: *Allerton Conference*, 2008, pp. 813–817, DOI: 10.1109/allerton.2008.4797641.

[301] Arezoo Rajabi, Rakesh B. Bobba, Mike Rosulek, Charles V. Wright, and Wu-chi Feng. “On the ImPracticality of Adversarial Perturbation for Image Privacy”. In: *Proceedings on Privacy Enhancing Technologies*, 2021, pp. 85–106, DOI: 10.2478/popets-2021-0006.

[302] Mehedi Hasan Raju, Dillon J Lohr, and Oleg V Komogortsev. “Evaluating Eye Movement Biometrics in Virtual Reality: A Comparative Analysis of VR Headset and High-End Eye-Tracker Collected Dataset”. In: 2024, DOI: https://doi.org/10.48550/arXiv.2405.03287.

[303] Vibhor Rastogi and Suman Nath. “Differentially private aggregation of distributed time-series with transformation and encryption”. In: *SIGMOD*, 2010, pp. 735–746, DOI: 10.1145/1807167.1807247.

[304] Vijay Ravi, Jinhan Wang, Jonathan Flint, and Abeer Alwan. “Enhancing accuracy and privacy in speech-based depression detection through speaker disentanglement”. In: *Comput. Speech Lang.*, 2024, DOI: 10.1016/j.csl.2023.101605.

[305] Dominick Reilly and Liyue Fan. “A Comparative Evaluation of Differentially Private Image Obfuscation”. In: *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, 2021, pp. 80–89, DOI: 10.1109/TPSISA52974.2021.00009.

[306] Davis Rempe, Tolga Birdal, Aaron Hertzmann, Jimei Yang, Srinath Sridhar, and Leonidas J. Guibas. “HuMoR: 3D Human Motion Model for Robust Pose Estimation”. In: *International Conference on Computer Vision (ICCV)*, 2021.

[307] Xiaojun Ren, Jiluan Fan, Ning Xu, Shaowei Wang, Changyu Dong, and Zikai Wen. “DPGazeSynth: Enhancing eye-tracking virtual reality privacy with differentially private data synthesis”. In: *Information Sciences*, 2024, p. 120720, DOI: 10.1016/j.ins.2024.120720.

[308] Kenneth Revett, Hamid Jahankhani, Sérgio Tenreiro de Magalhães, and Henrique Santos. “A survey of user authentication based on mouse dynamics”. In: *International Conference on Global e-Security*, 2008, pp. 210–219.

[309] Douglas A. Reynolds. “Speaker identification and verification using Gaussian mixture speaker models”. In: *Speech Communication*, 1995, pp. 91–108, DOI: 10.1016/0167-6393(95)00009-d.

[310] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. “Speaker Verification Using Adapted Gaussian Mixture Models”. In: *Digital Signal Processing*, 2000, pp. 19–41, DOI: 10.1006/dspr.1999.0361.

[311] Slobodan Ribaric, Aladdin Ariyaeeinia, and Nikola Pavesic. “De-identification for privacy protection in multimedia content: A survey”. In: *Signal Processing: Image Communication*, 2016, pp. 131–151, DOI: 10.1016/j.image.2016.05.020.

[312] Rokoko Electronics. *Rokoko Smart Suit*. https://www.rokoko.com/products/smartsuit-pro, Accessed: 03-03-2025.

[313] Pierre Rougé, Ali Moukadem, Alain Dieterlen, Antoine Boutet, and Carole Frindel. “Generalizable Features for Anonymizing Motion Signals Based on the Zeros of the Short-Time Fourier Transform”. In: *J. Signal Process. Syst.*, 2022, pp. 89–99, DOI: 10.1007/s11265-022-01798-9.

[314] Zhang Rui and Zheng Yan. “A survey on biometric authentication: Toward secure and privacy-preserving identification”. In: *IEEE access*, 2018, pp. 5994–6009.

[315] Michela Russo, Marianna Amboni, Paolo Barone, Maria Teresa Pellecchia, Maria Romano, Carlo Ricciardi, and Francesco Amato. “Identification of a Gait Pattern for Detecting Mild Cognitive Impairment in Parkinson’s Disease”. In: *Sensors*, 2023, p. 1985, DOI: 10.3390/s23041985.

[316] Chr. Ryf and A. Weymann. “The Neutral Zero Method — A Principle of Measuring Joint Function”. In: *Injury*, 1995, pp. 1–11, DOI: 10.1016/0020-1383(95)90116-7.

[317] Napa Sae-Bae and Nasir Memon. “A simple and effective method for online signature verification”. In: *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, 2013, pp. 1–12.

[318] Nazir Saleheen, Supriyo Chakraborty, Nasir Ali, Md Mahbubur Rahman, Syed Monowar Hossain, Rummana Bari, Eugene Buder, Mani Srivastava, and Santosh Kumar. “MSieve: Differential Behavioral Privacy in Time of Mobile Sensor Data”. In: *International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 706–717, DOI: 10.1145/2971648.2971753.

[319] Hosnieh Sattar, Katharina Krombholz, Gerard Pons-Moll, and Mario Fritz. “Body shape privacy in images: understanding privacy and preventing automatic shape extraction”. In: *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, 2020, pp. 411–428.

[320] Ben Saunders, Necati Cihan Camgoz, and Richard Bowden. "Anonysign: Novel Human Appearance Synthesis for Sign Language Video Anonymisation". In: *Automatic Face and Gesture Recognition*, 2021, pp. 1–8, DOI: 10.1109/FG52635.2021.9666984.

[321] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw. "BCI2000: a general-purpose brain-computer interface (BCI) system". In: *IEEE Transactions on Biomedical Engineering*, 2004, pp. 1034–1043, DOI: 10.1109/TBME.2004.827072.

[322] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. 2002.

[323] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *IEEE Computer Vision and Pattern Recognition*, 2015, DOI: 10.1109/cvpr.2015.7298682.

[324] Sefik Ilkin Serengil and Alper Ozpinar. "LightFace: A Hybrid Deep Face Recognition Framework". In: *IEEE Intelligent Systems and Applications Conference*, 2020.

[325] Abdur R. Shahid and Sajedul Talukder. "Evaluating Machine Learning Models for Handwriting Recognition-based Systems under Local Differential Privacy". In: *Innovations in Intelligent Systems and Applications Conference*, 2021, pp. 1–6, DOI: 10.1109/ASYU52992.2021.9598983.

[326] Ali Shahin Shamsabadi, Brij Mohan Lal Srivastava, Aurélien Bellet, Nathalie Vauquier, Emmanuel Vincent, Mohamed Maouche, Marc Tommasi, and Nicolas Papernot. "Differentially Private Speaker Anonymization". In: *Proceedings on Privacy Enhancing Technologies*, 2023, pp. 98–114.

[327] Shawn Shan, Emily Wenger, Jiayun Zhang, Huiying Li, Haitao Zheng, and Ben Y Zhao. "Fawkes: Protecting Privacy against Unauthorized Deep Learning Models". In: *29th USENIX security symposium (USENIX Security 20)*, 2020, pp. 1589–1604.

[328] Wenda Shao, Shiqing Luo, and Zhisheng Yan. "Cross-content User Authentication in Virtual Reality". In: *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, 2024, pp. 2098–2105, DOI: 10.1145/3636534.3696212.

[329] Dushyant Sharma, Francesco Nespoli, Rong Gong, and Patrick A. Naylor. "Canonical Voice Conversion and Dual-Channel Processing for Improved Voice Privacy of Speech Recognition Data". In: *2023 31st European Signal Processing Conference (EUSIPCO)*, 2023, pp. 66–70, DOI: 10.23919/EUSIPCO58844.2023.10289777.

[330] Michael Sherman, Gradeigh Clark, Yulong Yang, Shridatt Sugrim, Arttu Modig, Janne Lindqvist, Antti Oulasvirta, and Teemu Roos. "User-generated free-form gestures for authentication". In: *MobiSys*, 2014, pp. 176–189, DOI: 10.1145/2594368.2594375.

[331] Md Shopon, Sanjida Nasreen Tumpa, Yajurv Bhatia, K. N. Pavan Kumar, and Marina L. Gavrilova. "Biometric Systems De-Identification: Current Advancements and Future Directions". In: *Journal of Cybersecurity and Privacy*, 2021, pp. 470–495, DOI: 10.3390/jcp1030024.

[332] Noa Simhi and Galit Yovel. "Dissociating gait from static appearance: A virtual reality study of the role of dynamic identity signatures in person recognition". In: *Cognition*, 2020, p. 104445, DOI: 10.1016/j.cognition.2020.104445.

[333] Dipesh K. Singh, Gauri P. Prajapati, and Hemant A. Patil. “Voice Privacy Using Time-Scale and Pitch Modification”. In: *SN Comput. Sci.*, 2024, DOI: 10.1007/s42979-023-02549-8.

[334] Girijesh Singh, Palak Patel, Muhammad Asaduzzaman, and Garima Bajwa. “Selective EEG Signal Anonymization using Multi-Objective Autoencoders”. In: *2023 20th Annual International Conference on Privacy, Security and Trust (PST)*, 2023, pp. 1–7, DOI: 10.1109/PST58708.2023.10320167.

[335] David Snyder, Daniel Garcia-Romero, Gregory Sell, Daniel Povey, and Sanjeev Khudanpur. “X-Vectors: Robust DNN Embeddings for Speaker Recognition”. In: *Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5329–5333, DOI: 10.1109/icassp.2018.8461375.

[336] Cristina Soaz and Klaus Diepold. “Step Detection and Parameterization for Gait Assessment Using a Single Waist-Worn Accelerometer”. In: *TBME*, 2016, pp. 933–942, DOI: 10.1109/TBME.2015.2480296.

[337] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. “Learning structured output representation using deep conditional generative models”. In: *Advances in neural information processing systems*, 2015.

[338] Lal Srivastava, Brij Mohan, Nathalie Vauquier, Md Sahidullah, Aurelien Bellet, Marc Tommasi, and Emmanuel Vincent. “Evaluating Voice Conversion-Based Privacy Protection against Informed Attackers”. In: *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 2802–2806, DOI: 10.1109/ICASSP40776.2020.9053868.

[339] Ioanna-Ourania Stathopoulou and George A Tsihrintzis. “Emotion recognition from body movements and gestures”. In: *Intelligent interactive multimedia systems and services*. Springer, 2011, pp. 295–303.

[340] Julian Steil, Inken Hagestedt, Michael Xuelin Huang, and Andreas Bulling. “Privacy-Aware Eye Tracking Using Differential Privacy”. In: *ACM Eye Tracking Research & Applications*, 2019, pp. 1–9, DOI: 10.1145/3314111.3319915.

[341] Nathan J Stevenson, Karoliina Tapani, Leena Lauronen, and Sampsa Vanhatalo. “A dataset of neonatal EEG recordings with seizure annotations”. In: *Scientific data*, 2019, pp. 1–8.

[342] Tino Stöckel, Robert Jacksteit, Martin Behrens, Ralf Skripitz, Rainer Bader, and Anett Mau-Moeller. “The mental representation of the human gait in young and older adults”. In: *Frontiers in Psychology*, 2015, p. 943, DOI: 10.3389/fpsyg.2015.00943.

[343] Ariel Stolerman, Rebekah Overdorf, Sadia Afroz, and Rachel Greenstadt. *Classify, but verify: Breaking the closed-world assumption in stylometric authorship attribution*. Tech. rep. Drexel University, 2013, p. 17.

[344] Fahim Sufi, Seedahmed Mahmoud, and Ibrahim Khalil. “A new ECG obfuscation method: A joint feature extraction & corruption approach”. In: *Conference on Information Technology and Applications in Biomedicine*, 2008, pp. 334–337, DOI: 10.1109/itab.2008.4570644.

[345] Shravani Sur and VK Sinha. “Event-related potential: An overview”. In: *Industrial Psychiatry Journal*, 2009, p. 70, DOI: 10.4103/0972-6748.57865.

[346] Cees H. Taal, Richard C. Hendriks, Richard Heusdens, and Jesper Jensen. “A short-time objective intelligibility measure for time-frequency weighted noisy speech”. In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 4214–4217, DOI: 10.1109/ICASSP.2010.5495701.

[347] Takahiro Tamesue and Tetsuro Saeki. “Sound masking for achieving speech privacy with parametric acoustic array speaker”. In: *IEEE Conference on Soft Computing and Intelligent Systems*, 2014.

[348] Jimmy Tekli, Bechara al Bouna, Raphael Couturier, Gilbert Tekli, Zeinab al Zein, and Marc Kamradt. “A Framework for Evaluating Image Obfuscation under Deep Learning-Assisted Privacy Attacks”. In: *2019 17th International Conference on Privacy, Security and Trust (PST)*, 2019, pp. 1–10, DOI: 10.1109/PST47121.2019.8949040.

[349] Ömer Terlemez, Stefan Ulbrich, Christian Mandery, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. “Master Motor Map (MMM)—Framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots”. In: *2014 IEEE-RAS International Conference on Humanoid Robots*, 2014, pp. 894–901.

[350] Daksh Thapar, Aditya Nigam, and Chetan Arora. “Anonymizing Egocentric Videos”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 2300–2309, DOI: 10.1109/ICCV48922.2021.00232.

[351] Ngoc-Dung T. Tieu, Huy H. Nguyen, Fuming Fang, Junichi Yamagishi, and Isao Echizen. “An RGB Gait Anonymization Model for Low-Quality Silhouettes”. In: *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, 2019.

[352] Ngoc-Dung T. Tieu, Huy H. Nguyen, Hoang-Quoc Nguyen-Son, Junichi Yamagishi, and Isao Echizen. “An approach for gait anonymization using deep learning”. In: *IEEE Workshop on Information Forensics and Security*, 2017, pp. 1–6, DOI: 10.1109/WIFS.2017.8267657.

[353] Ngoc-Dung T. Tieu, Huy H. Nguyen, Hoang-Quoc Nguyen-Son, Junichi Yamagishi, and Isao Echizen. “Spatio-temporal generative adversarial network for gait anonymization”. In: *Journal of Information Security and Applications*, 2019.

[354] Ngoc-Dung T. Tieu, Junichi Yamagishi, and Isao Echizen. “Color Transfer to Anonymized Gait Images While Maintaining Anonymization”. In: *Asia-Pacific Signal and Information Processing Association Annual Symposium*, 2020, pp. 1406–1413.

[355] Julian Todt, Simon Hanisch, and Thorsten Strufe. “Fantômas: Understanding Face Anonymization Reversibility”. In: *Proc. Priv. Enhancing Technol.*, 2024, pp. 24–43, DOI: 10.56553/POPETS-2024-0105.

[356] N. Tomashenko, Brij Mohan Lal Srivastava, Xin Wang, Emmanuel Vincent, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Jose Patino, Jean-François Bonastre, Paul-Gauthier Noé, and Massimiliano Todisco. “Introducing the VoicePrivacy Initiative”. In: *Interspeech 2020*, 2020, DOI: 10.21437/interspeech.2020-1333.

[357] N. Tomashenko, Brij Mohan Lal Srivastava, Xin Wang, Emmanuel Vincent, Andreas Nautsch, Junichi Yamagishi, Nicholas Evans, Jose Patino, Jean-François Bonastre, Paul-Gauthier Noé, and Massimiliano Todisco. *Post-evaluation analysis for the VoicePrivacy 2020 Challenge: Using anonymized speech data to train attack models and ASR*. https://www.voiceprivacychallenge.org/docs/VoicePrivacy2020_post_evaluation.pdf.

[358] Natalia Tomashenko, Xin Wang, Xiaoxiao Miao, Hubert Nourtel, Pierre Champion, Massimiliano Todisco, Emmanuel Vincent, Nicholas Evans, Junichi Yamagishi, and Jean François Bonastre. "The VoicePrivacy 2022 Challenge Evaluation Plan". In: 2022, URL: https://arxiv.org/abs/2203.12468.

[359] Natalia Tomashenko, Xin Wang, Emmanuel Vincent, Jose Patino, Brij Mohan Lal Srivastava, Paul-Gauthier Noé, Andreas Nautsch, Nicholas Evans, Junichi Yamagishi, Benjamin O'Brien, Anaïs Chanclu, Jean-François Bonastre, Massimiliano Todisco, and Mohamed Maouche. "The VoicePrivacy 2020 Challenge: Results and findings". In: *Computer Speech & Language*, 2022, p. 101362, DOI: https://doi.org/10.1016/j.csl.2022.101362.

[360] Nikolaus F Troje. "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns". In: *Journal of Vision*, 2002, pp. 2–2, DOI: 10.1167/2.5.2.

[361] Nikolaus F Troje, Cord Westhoff, and Mikhail Lavrov. "Person identification from biological motion: Effects of structural and kinematic cues". In: *Perception & Psychophysics*, 2005, pp. 667–675, DOI: 10.3758/BF03193523.

[362] Charles Truong, Rémi Barrois-Müller, Thomas Moreau, Clément Provost, Aliénor Vienne-Jumeau, Albane Moreau, Pierre-Paul Vidal, Nicolas Vayatis, Stéphane Buffat, Alain Yelnik, Damien Ricard, and Laurent Oudre. "A Data Set for the Study of Human Locomotion with Inertial Measurements Units". In: *Image Processing On Line*, 2019, pp. 381–390, DOI: 10.5201/ipol.2019.265.

[363] Anthony Ngozichukwuka Uwaechia and Dzati Athiar Ramli. "A Comprehensive Survey on ECG Signals as New Biometric Modality for Human Authentication: Recent Advances and Future Challenges". In: *IEEE Access*, 2021, pp. 97760–97802, DOI: 10.1109/ACCESS.2021.3095248.

[364] Tavish Vaidya and Micah Sherr. "You Talk Too Much: Limiting Privacy Exposure Via Voice Input". In: *IEEE Security and Privacy Workshops*, 2019.

[365] Tamaya Van Criekinge, Ann Hallemans, Patricia Van De Walle, and Lizeth H. Sloot. "Age- and Sex-Related Differences in Trunk Kinematics during Walking in Able-Bodied Adults". In: *GeroScience*, 2023, pp. 2545–2559, DOI: 10.1007/s11357-023-01028-5.

[366] Laurens Van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE." In: *Journal of machine learning research*, 2008.

[367] Gabriele Vassallo, Tim Van hamme, Davy Preuveneers, and Wouter Joosen. "Privacy-Preserving Behavioral Authentication on Smartphones". In: *Workshop on Human-centered Sensing, Networking, and Systems*, 2017.

[368] *VRChat*. https://hello.vrchat.com/, Accessed: 2025-03-03.

[369] Changsheng Wan, Li Wang, and Vir V. Phoha. “A Survey on Gait Recognition”. In: *ACM Computing Surveys*, 2018, pp. 1–35, DOI: 10.1145/3230633.

[370] Han Wang, Shangyu Xie, and Yuan Hong. “VideoDP: A Flexible Platform for Video Analytics with Differential Privacy.” In: *Proceedings on Privacy Enhancing Technologies*, 2020.

[371] Hao Wang, Zhengquan Xu, Shan Jia, Ying Xia, and Xu Zhang. “Why current differential privacy schemes are inapplicable for correlated data publishing?” In: *World Wide Web*, 2021, pp. 1–23, DOI: 10.1007/s11280-020-00825-8.

[372] Shuo Wang, Ming Jiang, Xavier Morin Duchesne, Elizabeth A. Laugeson, Daniel P. Kennedy, Ralph Adolphs, and Qi Zhao. “Atypical Visual Saliency in Autism Spectrum Disorder Quantified through Model-Based Eye Tracking”. In: *Neuron*, 2015, pp. 604–616, DOI: 10.1016/j.neuron.2015.09.042.

[373] Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. “A Benchmark Dataset for SSVEP-Based Brain–Computer Interfaces”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2017, pp. 1746–1752, DOI: 10.1109/TNSRE.2016.2627556.

[374] Zhengyao Wen, Wenzhong Lin, Tao Wang, and Ge Xu. “Distract Your Attention: Multi-Head Cross Attention Network for Facial Expression Recognition”. In: *Biomimetics*, 2023, p. 199, DOI: 10.3390/biomimetics8020199.

[375] Ethan Wilson, Azim Ibragimov, Michael J. Proulx, Sai Deep Tetali, Kevin Butler, and Eakta Jain. “Privacy-Preserving Gaze Data Streaming in Immersive Interactive Virtual Reality: Robustness and User Experience”. In: *IEEE Transactions on Visualization and Computer Graphics*, 2024, pp. 2257–2268, DOI: 10.1109/TVCG.2024.3372032.

[376] Ge Wu, Sorin Siegler, Paul Allard, Chris Kirtley, Alberto Leardini, Dieter Rosenbaum, Mike Whittle, Darryl D D’Lima, Luca Cristofolini, Hartmut Witte, Oskar Schmid, and Ian Stokes. “ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine”. In: *Journal of Biomechanics*, 2002, pp. 543–548, DOI: 10.1016/S0021-9290(01)00222-6.

[377] Ge Wu, Frans C.T. van der Helm, H.E.J. (DirkJan) Veeger, Mohsen Makhsous, Peter Van Roy, Carolyn Anglin, Jochem Nagels, Andrew R. Karduna, Kevin McQuade, Xuguang Wang, Frederick W. Werner, and Bryan Buchholz. “ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: shoulder, elbow, wrist and hand”. In: *Journal of Biomechanics*, 2005, pp. 981–992, DOI: 10.1016/j.jbiomech.2004.05.042.

[378] Shun-Chi Wu, Peng-Tzu Chen, A. Lee Swindlehurst, and Pei-Lun Hung. “Cancelable Biometric Recognition With ECGs: Subspace-Based Approaches”. In: *IEEE Transactions on Information Forensics and Security*, 2019, pp. 1323–1336, DOI: 10.1109/tifs.2018.2876838.

[379] Danny Wyatt, Tanzeem Choudhury, and Jeff Bilmes. “Conversation detection and speaker segmentation in privacy-sensitive situated speech data”. In: *Interspeech 2007*, 2007, pp. 586–589, DOI: 10.21437/Interspeech.2007-256.

[380] Zhaoyang Xia, Yuxiao Chen, Qilong Zhangli, Matt Huenerfauth, Carol Neidle, and Dimitris Metaxas. “Sign Language Video Anonymization”. In: *Workshop on the Representation and Processing of Sign Languages*, 2022.

[381] Shilin Xiao, Xiaoyu Ji, Chen Yan, Zhicong Zheng, and Wenyuan Xu. “MicPro: Microphone-based Voice Privacy Protection”. In: *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 2023, pp. 1302–1316, DOI: 10.1145/3576915.3616616.

[382] Sherif Yacoub, Steve Simske, Xiaofan Lin, and John Burns. “Recognition of emotions in interactive voice response systems”. In: *EUROSPEECH*, 2003.

[383] Neil Yager and Ted Dunstone. “The Biometric Menagerie”. In: *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, pp. 220–230, DOI: 10.1109/TPAMI.2008.291.

[384] Roman V. Yampolskiy and Venu Govindaraju. “Taxonomy of Behavioural Biometrics”. In: *Behavioral Biometrics for Human Identification*. IGI Global, 2010, pp. 1–43. DOI: 10.4018/978-1-60566-725-6.ch001.

[385] Qing Yang, Tao Wang, Ning Su, Shifu Xiao, and Zoi Kapoula. “Specific saccade deficits in patients with Alzheimer’s disease at mild to moderate stage and in patients with amnestic mild cognitive impairment”. In: *AGE*, 2012, pp. 1287–1298, DOI: 10.1007/s11357-012-9420-z.

[386] Yang Yang, Yury Kartynnik, Yunpeng Li, Jiuqiang Tang, Xing Li, George Sung, and Matthias Grundmann. “STREAMVC: Real-Time Low-Latency Voice Conversion”. In: *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 11016–11020, DOI: 10.1109/ICASSP48485.2024.10446863.

[387] Yulong Yang, Gradeigh D Clark, Janne Lindqvist, and Antti Oulasvirta. “Free-form gesture authentication in the wild”. In: *Conference on Human Factors in Computing Systems*, 2016, pp. 3722–3735.

[388] Jixun Yao, Qing Wang, Pengcheng Guo, Ziqian Ning, and Lei Xie. “Distinctive and Natural Speaker Anonymization via Singular Value Transformation-Assisted Matrix”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024, pp. 2944–2956, DOI: 10.1109/TASLP.2024.3407600.

[389] Jixun Yao, Qing Wang, Yi Lei, Pengcheng Guo, Lei Xie, Namin Wang, and Jie Liu. “Distinguishable Speaker Anonymization Based on Formant and Fundamental Frequency Scaling”. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5, DOI: 10.1109/ICASSP49357.2023.10095120.

[390] Xin Yao and Senquan An. “DP-VoicePub: Differential Privacy-based Voice Publication”. In: *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2023, pp. 1–5, DOI: 10.1109/ISCAS46773.2023.10182113.

[391] Yue Yao, Josephine Plested, Tom Gedeon, Yuchi Liu, and Zhengjie Wang. “Improved Techniques for Building EEG Feature Filters”. In: *IEEE Conference on Neural Networks*, 2019, pp. 1–6, DOI: 10.1109/ijcnn.2019.8852302.

[392] Tal Yarkoni, Christopher J Markiewicz, Alejandro de la Vega, Krzysztof J Gorgolewski, Taylor Salo, Yaroslav O Halchenko, Quinten McNamara, Krista DeStasio, Jean-Baptiste Poline, Dmitry Petrov, et al. "PyBIDS: Python tools for BIDS datasets". In: *Journal of open source software*, 2024, p. 1294, DOI: 10.5281/ZENODO.11244297.

[393] Tal Yarkoni, Christopher J. Markiewicz, Alejandro de la Vega, Krzysztof J. Gorgolewski, Yaroslav O. Halchenko, Taylor Salo, Quinten McNamara, Krista DeStasio, Jean-Baptiste Poline, Dmitry Petrov, Valérie Hayot-Sasson, Dylan M. Nielson, Johan Carlin, Gregory Kiar, Kirstie Whitaker, Adina Wagner, Elizabeth DuPre, Stefan Appelhoff, Alexander Ivanov, Johannes Wennberg, Lee S. Tirrell, Oscar Esteban, Mainak Jas, Michael Hanke, Russell Poldrack, Chris Holdgraf, Isla Staden, Ariel Rokem, Bertrand Thirion, Chadwick Boulay, Dave F. Kleinschmidt, Erin W Dickie, John A. Lee, Matteo Visconti di Oleggio Castello, Michael Philipp Notter, Pauline Roca, and Ross Blair. "bids-standard/pybids: 0.9.3". In: 2019, DOI: 10.5281/zenodo.3363985.

[394] In-Chul Yoo, Keonnyeong Lee, Seonggyun Leem, Hyunwoo Oh, Bonggu Ko, and Dongsuk Yook. "Speaker Anonymization for Personal Information Protection Using Voice Conversion Techniques". In: *IEEE Access*, 2020, pp. 198637–198645, DOI: 10.1109/ACCESS.2020.3035416.

[395] Galit Yovel and Alice J O'Toole. "Recognizing people in motion". In: *Trends in cognitive sciences*, 2016, pp. 383–395, DOI: 10.1016/j.tics.2016.02.005.

[396] Shiqi Yu, Tieniu Tan, Kaiqi Huang, Kui Jia, and Xinyu Wu. "A Study on Gait-Based Gender Classification". In: *IEEE Transactions on Image Processing*, 2009, pp. 1905–1910, DOI: 10.1109/TIP.2009.2020535.

[397] Ruibin Yuan, Yuxuan Wu, Jacob Li, and Jaxter Kim. "DeID-VC: Speaker De-identification via Zero-shot Pseudo Voice Conversion". In: *Interspeech 2022*, 2022, pp. 2593–2597, DOI: 10.21437/Interspeech.2022-11036.

[398] Emna Kalai Zaghouani, Adel Benzina, and Rabah Attia. "ECG based authentication for e-healthcare systems: Towards a secured ECG features transmission". In: *IEEE Wireless Communications and Mobile Computing Conference*, 2017, pp. 1777–1783, DOI: 10.1109/iwcmc.2017.7986553.

[399] Emna Kalai Zaghouani, Adel Benzina, and Rabah Attia. "ECG biometrie template protection based on secure sketch scheme". In: *IEEE Software, Telecommunications and Computer Networks*, 2017, pp. 1–5, DOI: 10.23919/softcom.2017.8115526.

[400] Mohammad-Reza Zare-Mirakabad, Fatemeh Kaveh-Yazdy, and Mohammad Tahmasebi. "Privacy preservation by k-anonymizing Ngrams of time series". In: *ISC Conference on Information Security and Cryptology*, 2013, pp. 1–6, DOI: 10.1109/ISCISC.2013.6767335.

[401] Gao Zhang, Zhiwei Guan, Guozhong Dai, and Xiangshi Ren. "A comparison of four interaction modes for CAD systems". In: *APCHI*, 1998, pp. 82–87, DOI: 10.1109/APCHI.1998.704160.

[402] Guanglin Zhang, Sifan Ni, and Ping Zhao. "Enhancing Privacy Preservation in Speech Data Publishing". In: *Internet of Things Journal*, 2020, pp. 7357–7367, DOI: 10.1109/jiot.2020.2983228.

[403] Jie Zhang and Robert B. Fisher. "3D Visual passcode: Speech-driven 3D facial dynamics for behaviometrics". In: *Signal Process.*, 2019, pp. 164–177, DOI: 10.1016/j.sigpro.2019.02.025.

[404] Ni Zhang and Yoshinori Yaginuma. "A privacy-preserving and language-independent speaking detecting and speaker diarization approach for spontaneous conversation using microphones". In: *2012 IEEE 11th International Conference on Signal Processing*, 2012, pp. 499–502, DOI: 10.1109/icosp.2012.6491534.

[405] Wei Zhang, Xianpeng Ji, Keyu Chen, Yu Ding, and Changjie Fan. "Learning a Facial Expression Embedding Disentangled from Identity". In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 6755–6764, DOI: 10.1109/cvpr46437.2021.00669.

[406] Zengqun Zhao, Qingshan Liu, and Feng Zhou. "Robust Lightweight Facial Expression Recognition Network with Label Distribution Training". In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 3510–3519, DOI: 10.1609/aaai.v35i4.16465.

[407] Jianwei Zheng, Jianming Zhang, Sidy Danioko, Hai Yao, Hangyuan Guo, and Cyril Rakovski. "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients". In: *Scientific Data*, 2020, DOI: 10.1038/s41597-020-0386-x.

[408] Yu Zhong and Yunbin Deng. "A Survey on Keystroke Dynamics Biometrics: Approaches, Advances, and Evaluations". In: *Recent Advances in User Authentication Using Keystroke Dynamics Biometrics*. Science Gate Publishing P.C., Jan. 2015, pp. 1–22. DOI: 10.15579/gcsr.vol2.ch1.

[409] Yuhan Zhou, Robbin Romijnders, Clint Hansen, Jos van Campen, Walter Maetzler, Tibor Hortobágyi, and Claudine JC Lamoth. "The detection of age groups by dynamic gait outcomes using machine learning approaches". In: *Scientific Reports*, 2020, pp. 1–12, DOI: 10.1038/s41598-020-61423-2.

[410] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, and Jie Zhou. "WebFace260M: A Benchmark Unveiling the Power of Million-scale Deep Face Recognition". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[411] Rick van der Zwan, C MacHatch, Desiree Kozlowski, NF Troje, O Blanke, and Anna Brooks. "Gender bending: auditory cues affect visual judgements of gender in biological motion displays". In: *Experimental Brain Research*, 2009, pp. 373–382, DOI: 10.1007/s00221-009-1800-y.

A. Behavioral Data Anonymizations

A.1. Voice

Voice processing and analysis [21] have long been performed and hence a large set of specific terminology exists to describe it. The sound of the human voice is created by the Larynx and then travels via the vocal tract, which transforms and filters the sound before it leaves the mouth. Due to its approximate tube shape, the vocal tract produces resonances of the sound which are dependent on the length of the vocal tract. A Phoneme is the smallest unit of sound that distinguishes one word from another and an utterance is a unit of speech between two clear pauses. The log-spectrum is an important representation of sound as it is closer to human perception. By using a domain transformation (fast Fourier transform (FFT) or cosine) on the log-spectrum we get the cepstrum (see Figure A.1). The cepstrum is useful because it allows easy estimation of the fundamental frequency (f0) of the signal. The perceived fundamental frequency by humans is known as pitch. A widely used scale to transform the fundamental frequency to the pitch is the Mel scale. Using the Mel scale the cepstrum can be sampled at frequencies with the same perceived distance using weighted sums. Applying an FFT on those sums gives the Mel-frequency cepstral coefficients (MFCC). The MFCCs are an approximate quantification of the signal spectrum that focuses on the macrostructure of the signal.

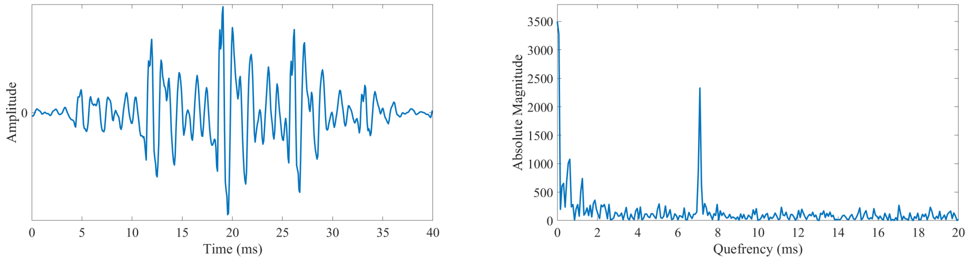


Figure A.1.: A windowed speech segment (left) and its corresponding Cepstrum (right), Source: `https://wiki.aalto.fi/display/ITSP/Cepstrum+and+MFCC`.

The following gives a short overview of the field of speaker recognition (i.e., identification) which aims to establish the identity of a speaker. Gaussian mixture models [309] (GMM) represent speakers as the distribution of their feature vectors. The feature vectors are extracted from the speech (most often represented as MFCC) of the speaker and then modeled as Gaussian mixture density. A GMM assumes that the data points are generated by a finite number of Gaussian distributions with unknown parameters. Each feature vector is represented as a linear combination of Gaussian densities. A universal background model (UBM) is a GMM that models a wide variety of non-target speakers, representing possible impostors. The means of the UBM are then adjusted to the target speaker by using a maximum a posteriori adaption [310] resulting in a GMM for the target speaker. The benefit of this approach is that the Gaussians used to model the target speaker are the same as in the UBM. For the classification of a speaker, the log-likelihood of the target speaker GMM is compared to that of the UBM to determine if the speaker should be accepted. An alternative to the log-likelihood approach is to get a GMM for each speaker recording through a maximum a posteriori probability (MAP) adaptation of the UBM and then map these GMM to a new feature vector, called Supervector [44]. Supervectors can be classified using traditional methods like support vector machines. An extension of Supervectors is the total variability (TV) [70] approach. This maps the Supervectors to a low-dimensional space that models both the speaker and the channel variability. The resulting vector is called i-vector and is the de facto state-of-art in speaker identification. An alternative to i-vectors are x-vectors [335] which are extracted for each utterance via a deep neural network (DNN).

A.1.0.1. Utility

The main usage of voice recordings is the transmission of information between humans, however, in recent years voice also became an important input modality for computer systems [292]. In both cases, it is important that the content of the speech is intelligible for the intended listeners. But also the mere detection of speech in audio samples can be useful, for example for crowd detection [57]. Further, voices uniquely identify their speaker, making them suitable both for authentication and recognition purposes [314].

A.1.0.2. Threat Space

The privacy threats for human voices range from the identification of individuals, over the inference of private attributes, to identity theft via fake recordings. The identification of individuals via their voice has long been apparent to humans. But voices convey more information than just identity, they also allow us to infer attributes such as gender [84], or emotional state [382]. Further, modern speech synthesis methods allow the creation of fake voice recordings for a target speaker, enabling identity theft or the circumvention of speaker authentication systems. Unlike other behavioral biometric traits, voice and its resulting speech can also carry a semantic meaning, which can be sensitive to privacy.

A.1.0.3. Additional Privacy Goal

Voice has speech blurring as an additional privacy goal, which aims at destroying the intelligibility of the speech to protect its semantic content from unintentional listeners.

A.1.0.4. Anonymization Techniques

We now present the surveyed anoymization techniques that deal with protecting human voices.

Random Perturbation Parthasarathi et al. [270] extend their feature removal methods [271] by additionally shuffling the voice blocks for adding randomness. Mtibaa et al. [248] propose a template protection scheme that relies on shuffling the feature vector of a GMM-UBM speaker identification system.

Noise Injection Tamesue et al. [347] propose a very simple method to make speech unintelligible by simply playing pink noise between 180 and 5630 Hz with various dBs. Ma et al. [204] also try to make speech unintelligible but focus on smartphone recordings. Their device creates two ultrasound waves whose interaction creates random low frequency waves that noise the microphone of a smartphone but cannot be heard by humans. In their evaluation, they found that they can block smartphone recordings up to 5 meters, depending on the type of smartphone. Hashimoto et al. [120] propose a system to preserve speaker privacy in physical spaces. The core idea is to add white noise to prevent recordings of speakers from being used for identity theft. They conclude that preventing speaker identification is possible (equal error rate (EER) from 2% to 17%) while at the same time keeping the intelligibility of the speech at a high level (short-time objective intelligibility [346] from 1 to 0.9).

Ohshio et al. [265] train multiple so-called babble maskers from pre-recorded speakers by segmenting the speech and then averaging the segments. When a speaker should be de-identified the babble masker is selected based on the fundamental frequency and the pitch of the person. Vaidya et al. [364] proposes to add random noise to four features: pitch, tempo, pause, and MFCC. We found the descriptions of their approach to be rather short. Sharma et al. [329] use a self attention channel combinator to add noise to voice signals.

Two methods have been proposed that rely on differential privacy for noise injection. Hamm et al. [113] propose a differential private min-max filter. The min-max filter minimizes the privacy risk while maximizing utility risk with a given utility and private task. The differential privacy is achieved by adding noise either in front of the filter or after the filter. Han et al. [114] rely on X-vectors as speaker representation and formally define voice-indistinguishably a privacy metric using differential privacy. As a measurement of similarity between x-vectors, the angular distance is used and the overall scheme gives an upper limit of this distance until which two x-vectors cannot be distinguished.

Feature Removal Parthasarathi et al. [272] propose three feature removal methods for privacy-aware speaker change detection. Adaptive filtering assumes that the excitation source is independent of the vocal tract response. They perform short-term linear prediction analysis to estimate an all-pole model [185](representing the vocal tract), a residual (representing the excitation source), and the gain. Then the residual is used to estimate its real cepstrum. Their second method is to remove all subbands except the one from 1.5 kHz to 2.5 kHz and from 3.5 kHz to 4.5 kHz. They represent the two subbands as MFCC coefficients and log-energy from a single filter. Their last method only uses the spectral slope of the speaker represented as cepstral coefficients. In another work [271] Parthasarathi et al. also propose similar feature removal methods for speaker diarisation using the real cepstrum and MFCC as features. Their analysis finds that MFCC works better than real cepstrum. Agarwal et al. [8] propose a similar scheme. They first transform the segmented speech signals into the frequency domain, then select the n most important peaks and interpolate a new signal before transforming it back into the speech domain.

Wyatt et al. [379] propose a feature removal method for speaker segmentation and conversation detection. They split the audio into segments and save for each the non-initial maximum autocorrelation peak, the total number of autocorrelation peaks, the relative spectral entropy, and the energy of the frame. Zhang et al. [404] use the same features as proposed by Wyatt et al. except for the energy of the frame and then use an HMM to perform the conversation detection. An evaluation of privacy is missing in both works.

Ditthapron et al. [76] have investigated how speech from non-target speakers can be removed in a speech assessment scenario. To separate the speakers they first extract speaker representations from the MFCC of the speech via an encoder. The speaker representation is then concatenated with the original MFCC before all but the target speakers are filtered out in the speaker matching network. We are missing a convincing evaluation of privacy.

Nelus et al. [255] propose to train a DNN via adversarial learning to extract features from a speaker that allow gender recognition but not speaker identification. Their evaluation shows a drop in identification from 61% to 26% while the gender recognition only drops by 1%. They also proposed a similar system [256] which removes speaker identities from urban sound recordings. Cohen-Hadria et al. [57] also use a neural network and use it to extract the voices from recordings that consist of both background and voice noise in which the voices should be anonymized. They remove attributes with two methods. The first method simply low-pass filters the voice at 250 Hz. The second method extracts the MFCC from the voice and then uses the first 5 components to create a new voice. In the end, the blurred speech is recombined with the background noise. Evaluating with a speaker identification system they were able to reduce the identification down to 29% from 43%.

Discrete Conversion For discrete conversion, we found multiple template protection schemes.

Pathak et al. [274] present a hashing algorithm to protect voice data for authentication purposes. The supervector of a speaker is gained by performing the MAP adaptation of a universal background model for each utterance of the speaker and concatenating the means of the adapted model. The locality sensitive hashing is then performed with the supervector which transforms it into a low dimensional space, which is referred to as a bucket. This operation is an approximation of the nearest neighbors algorithm allowing the comparison of buckets to authenticate the individual.

Portelo et al. [290, 291] propose a template protection scheme based on secure binary embeddings. The authors use a speaker identification system that uses supervectors and i-vectors to represent the features of a speaker's voice. The feature vectors are then encoded with secure binary embeddings which have the property that if the Euclidean distance of the two vectors is below a certain threshold then the hamming distance of the resulting hashes is proportional to the Euclidean distance. This allows the comparison of the encoded vectors by using a support vector machine (SVM) with a hamming distance based kernel.

Billeb et al. [34] propose a template protection scheme that is based on fuzzy commitment. They first extract the frequency spectrum via an FFT and then extract features from the magnitude spectrum. Then the MAP adaptation of a GMM-UBM speaker identification system is applied and additional statistics are extracted. The template is then stored as a combination of error-correcting code and hash algorithm.

Continuous Conversion Most voice anonymization techniques fall into the category of continuous conversion, since they attempt to create an anonymized speech recording. We have found the following techniques.

Speaker transformation is the process of manipulating the voice characteristics of a speaker (not the linguistic features) to make the voice sound like a target speaker. A target speaker can be either a specific natural speaker or a synthetic speaker. For the synthetic speaker, either an existing speaker is used or a new one is generated, for example by averaging multiple speakers into one. The general approach of speaker transformation is that the voice characteristics of the source speaker are extracted and then transformed to match the target speaker. In the last step, the new speaker is synthesized. The following methods perform speaker transformation.

Jin et al. [150] evaluate four methods for speaker transformation for identity protection. Their base method uses a GMM-mapping based speaker transformation system to transfer speakers to a target synthetic voice called kal-diphone. Further, they test duration transformation in which the length of utterances of the source speaker is scaled to match the ones of the target speaker. Lastly, they try an extrapolated transformation in which they use the linear mapping of the source to the target to extrapolate beyond the target. Pobar et al. [285] also use a speaker transformation system based on GMM mapping but combine it with a harmonic stochastic model. The system is trained on a set of speakers to learn the transformation functions. Instead of retraining the system for a new speaker one of the existing transformation functions is applied. This removes the need for a parallel corpus for the speakers that should be protected. The target speaker is a synthetic speaker which reduces the identification accuracy from 97% down to 9%.

Justin et al. [155] investigate the intelligibility of transformed speakers. They test with a diphone speech synthesis system and an HMM-based speech synthesis system to transform speakers into a synthetic speaker. They performed a survey with human listeners to evaluate the intelligibility of the protected speakers, measuring the word error rate. Abou-Zleikha et al. [4] do not propose a speaker transformation method themselves but explore how to select a target speaker to achieve the lowest identification rate and have good results when the speaker is transformed back to the source speaker. They formulate this as an optimization problem and measure the distance between two speakers with a confusion factor, for which they evaluate entropy and Gini index as metrics. Pribil et al. [295] propose a speaker de-identification method that relies on modifying several features of the source speaker. In the first step, the prosodic and spectral features are extracted from the source speaker. They then modify the features to make the speaker sound older, younger, more female, and more male by using manually defined transformation functions and feature differences for each class. After the features are modified the de-identified speaker is synthesized.

Bahamanienezhad et al. [23] have developed a speaker transformation method that uses a convolutional encoder/decoder network. They, first extract spectral features and excitation features (f0) from the source speaker. The spectral features are then mapped via the encoder/decoder framework to a target speaker. The resulting speech is fused either via taking the average or via a gender-based average to create an average speaker. From the excitation features, only the fundamental frequency is transformed via linear transformation, the remaining features stay the same. Both spectral and excitation features are used to synthesize the de-identified speaker.

Fang et al. [89] use a similar averaging approach but rely on x-vectors. They extract the x-vector of a speaker and then use a set of random x-vectors of unrelated speakers to calculate a mean x-vector. In their evaluation, they demonstrate EER up to 34% for their anonymization. Mawalim et al. [221] propose to improve the system by Fang et al. by scaling the f0 frequency either up or down, increasing the length of the speech utterances by 1.2, and using singular value modification for the combination of the x-vectors. Their EER improved up to 54%. Further improved was this system by Prajapati et al. [293] who added a CycleGAN to modify the speakers. Cheng et al. [52] propose another speaker transformation which uses one encoder for content and one for the speaker identity and then recombines them in a single decoder into the anonymized utterance. Panarielle et al. [266] use neural audio codecs (NAC) for the speaker transformation. Similar to other speaker transformation techniques, they independently encode the content and the speaker identity and then combine them using transformer models before decode them using the NAC decoder.

As more speaker transformations appear, some of them focus on specific subproblems of speaker transformation. Miao et al. [235] developed a speaker transformation that is language independent. The architecture of the system is based on the B1 baseline of the VoicePrivacy [358] challenge and they are able to show that their system works on both English and Mandarin speaker datasets. Hintz et al. [125] investigate how to anonymize stuttering speakers using a GAN to preserve the pathology of the stuttering intact while removing the speaker identity. Yang et al. [386] have developed a low-latency speaker transformation technique. Yao et al. [389] attempt to improve the distinctiveness of anonymized speakers by scaling the formant and pitch information. Meyer et al. [234] propose a speaker transformation method that preserves the prosody of the speaker. Nespoli et al. [257] propose to use two speaker transformation systems in a row to achieve better anonymization results.

Several papers investigate how the target speaker for speaker transformation can be either selected or created using the original speaker as a starting point. Chang et al. [47] and Meyer et al. [233] investigate different averaging strategies. Yuan et al. [397] train an autoencoder and use it for synthetic data generation to generate random speakers. Lv et al. [203] use autoencoders to obtain a latent representation of the speaker and then select similar latent representations from a pool using k-means. Yao et al. [388] encode the speaker as a matrix. This matrix is then decomposed using singular value decomposition (SVD) into eigenvectors and a matrix that stores the importance of each eigenvector. They then use a logarithmic transformation to make the importance values more similar before reconstructing the speaker identity matrix. Miao et al. [236] extend their method [235] by removing the speaker pool and using an adversarial perturbation to transform the speaker vector. Perero-Codosera et al. [280] also propose an approach using an adversarial perturbation for anonymizing the original speaker X-vector, and Yao et al. [390] propose removing random dimensions of an X-vector to create a new speaker identity.

Adversarial Perturbation: In recent years, the technique of adversarial perturbation has become popular. The general idea is that the anonymization is performed by a machine learning system which is trained with two losses. One loss is for the privacy attribute to be protected and should be minimized while the other loss is for the desired utility and should be maximized.

Cheng et al. [48] propose VoiceCloak, which trains a convolutional perturbation injector to take the room impulse response and the original voice signal as input and outputs an anonymized voice. Deng et al. [72] present V-Cloak, which uses a convolutional autoencoder trained to minimize identification while preserving the timbre and intelligibility of the utterance high. Unique to the system is that it feeds the down-sampled speech into most of the layers of the autoencoder. In their evaluation, they can show that their system can be used for real-time anonymization and performs better or as good as other state-of-the-art voice anonymizations.

Chouchane et al. [56] address fairness concerns in speaker verification. They use adversarial training to create a speaker verification system that produces speaker embeddings that can still be used for speaker verification but no longer work for sex recognition. Xiao et al. [381] have developed a microphone module that anonymizes the speaker by adding an adversarial perturbation to the sound signal encoded by a generic code excitation linear prediction (CELP) codec. An interesting distinction from other adversarial perturbations is that they use a genetic algorithm to find the adversarial perturbation rather than gradient descent. They also show that their approach adds very little latency overhead. Ravi et al. [304] developed an adversarial perturbation for the utility goal of depression detection in speakers.

Ali et al. [10] also propose an autoencoder to anonymize at the network edge specifically for the input of voice assistants. Their idea is to extract privacy friendly features by training classifiers on the latent code of the voice samples. They use the trained classifier to perform gradient reversal on the encoder to unlearn the features learned for identity, gender, and language. Yoo et al. [394] use a CycleGAN for speaker anonymization which uses a variational autoencoder as its generator. They train against a DNN speaker recognition system as the discriminator.

Frequency warping is a technique that is similar to speaker transformation, the main difference is that frequency warping focuses on transforming the frequency spectrum of a speaker and usually does not try to transform the source into a specific target speaker. It is mostly used for identity and gender protection. A common goal of frequency warping is vocal tract length normalization in which the resonances that are specific to an individual's vocal tract length should be removed or altered.

Faundez-Zanuy et al. [90] explore two approaches for gender protection: Phase vocoder and vocal tract length normalization. The vocoder approach detects peaks in the voice signal. For each peak, a bin is defined and compared to its two neighbors to define a region of influence. Then the peak and its region of influence are shifted by a peak specific frequency. For both genders they can reduce gender recognition to chance level, however the identity recognition is also close to chance level. Valdivielso et al. [1] present a speaker protection approach that transforms the pitch and the frequency axis. Lopez-Otero et al. [197] rely on frequency warping and amplitude scaling for speaker protection in the context of depression detection. They implement both operations as an affine transformation in the cepstral domain and manually define piece-wise linear transformation functions. They demonstrate an increase of the EER from 9.7% to up to 44% for the speaker identification, while the depression detection stays similar to the clear data.

Magarinos et al. [205] also rely on frequency and amplitude warping for speaker protection. First, they extract the cepstral voice vectors from the speaker and then convert them into a discrete spectrum. Then dynamic frequency warping (DFW) is applied to map the source spectrum bins to the target spectrum. As multiple source bins can have the same target bin, all source bins that map to the same target bin are averaged. Additionally to the frequency and amplitude warping the fundamental frequency is adjusted regarding its mean and variance. They demonstrate an identification reduction from 99% to 4%.

Aloufi et al. [12] try to hide the emotional state of speakers before their speech is sent to a voice-based cloud service. They first extract the fundamental frequency, spectral envelope, and aperiodicity. The features are then transformed via a CycleGAN from emotional speech to neutral speech. Their framework has three modi, the first removes private attributes, the second removes the identity, and the third removes the intelligibility of the speech. Specific to this approach is that two separate encoders are used, one to encode the speech and one to encode the speaker. Their results for hiding the emotional state show a reduction from over 70% to about 20% and for hiding sex a reduction from up to 99% to the chance level of 50%.

Srivastava et al. [338] evaluate multiple speaker protection methods against an informed attacker. They work with three attacker models: An ignorant attack that is not aware that the voice data is de-identified, a semi-informed attacker that knows that the data is de-identified, and an informed attacker that knows the de-identification method and its parameters. The first method is a vocal tract length normalization approach. The speaker is represented as a set of centroid spectra. The algorithm then calculates the closest path between the source set and the target set to get the parameters for the warping. The second method uses a neural net encoder/decoder approach to transform the speaker. They found large differences for the different attacker models, while the ignorant attacker can achieve EER of up to 50% the informed attacker only achieves 11% as its highest EER. This finding highlights how important strong attacker models are for the evaluation of anonymization techniques.

Patino et al. [275] pseudonymize speakers by transforming their McAdam coefficients. In the first step linear predictive coding (LPC) is applied to an input speech frame. The coefficients of the LPC are then transformed into poles and the poles which have a nonzero imaginary part are shifted according to the angle between the real and imaginary part of the pole. Their evaluation shows that this approach performs well against an ignorant attack that is not aware of the anonymization increasing ERR from 3% to 26% while an informed attacker still achieves 5% ERR. Gupta et al. [109] further, improve on transforming the McAdams coefficients by not only changing the angle of the complex poles but also modifying their radius.

Mawalim et al. [222] propose two frequency modifications for voice anonymization. Their first technique segments the speech signal and then resamples the segments to raise or lower the pitch. They then use a Hann window function to combine the segments into the speech signal. Their second technique uses a different recombination technique by recombining the overlapping segments using phase propagation. Gaznepoglu et al. [100] modify the B1 baseline of the VoicePrivacy challenge [358] to produce better anonymized fundamental frequencies by first extracting them from X-vectors and then using a mask to anonymize them.

Continuous Conversion + Random Perturbation Canuto et al. [46] proposes a new method for template protection in which the feature vector is shuffled via a randomized sum. For each feature vector, the elements are shuffled based on a secret key. Two random vectors of the same length are derived from the key. These vectors give the position of the attributes that should be summed. The reorganized feature vector is summed up with the vectors resulting when the position vectors are applied to the original feature vector.

Prajapati et al. [294, 333] first use a regular voice conversion system and then perturb the speed of the speech sequence by changing the length of the sequence. They also adjust the tempo of the sequence by cutting the sequence into segments and making them randomly shorter or longer. They recombine the segments by using a overlap-add method. Their evaluation shows that the speed perturbation makes the anonymization stronger.

Continuous Conversion + Noise Injection Kondo et al. [167, 168] create so-called babble maskers by segmenting speech into ten second segments and then averaging them into babble maskers. Besides speaker-dependent maskers, they also create gender-based babble maskers based on multiple speakers of the same gender. The babble masker is then applied to the recording of the speaker. Qian et al. [297] present a method to sanitize speech before it is sent to the server of a virtual assistant. Their main method is to perform vocal tract length normalization via a compound frequency warping function consisting of a bilinear and a quadratic function to avoid re-identification attacks. Additionally, they add Laplace noise after the warping function to make the anonymization more robust. For the result, they claim to achieve differential privacy. In a follow up work [298] the same authors further investigate the security of their scheme. Srivastava et al. [338] also investigate the security of the scheme with stronger attackers.

Shmsabadi et al. [326] aim to provide theoretical privacy guarantees for speaker transformation. They do this by adding differential privacy to the pitch and context features used in speaker transformation. Both features are encoded by a specific autoencoder network, which transfers them into their latent space. For the pitch they then add Laplace noise and then perform a clipping of the latent vector values before decoding back to pitch space. For the context features they first normalize the latent vector and then add the Laplace noise before normalizing again and then decoding back. Due to the correlations between speech segments, it is unclear whether the differential privacy guarantees hold.

A.1.0.5. Evaluations

Most of the reviewed works evaluate the quality of the de-identification by comparing the recognition rates of attributes or identities on unmodified and de-identified data. The recognition is done via machine learning models or human listeners. As metrics to measure the recognition rate the papers mostly rely on the equal error rate (EER), false positive rate (FPR), false negative rate (FNR), recall, precision, and F1 score. Abou-Zleikha et al. [4] also use entropy and the Gini index to evaluate the de-identification performance. We believe that the prevalence of EER shows that the underlying scenario focuses on speaker verification scenarios, but we believe that speaker identification is a more appropriate scenario for evaluating speaker anonymization.

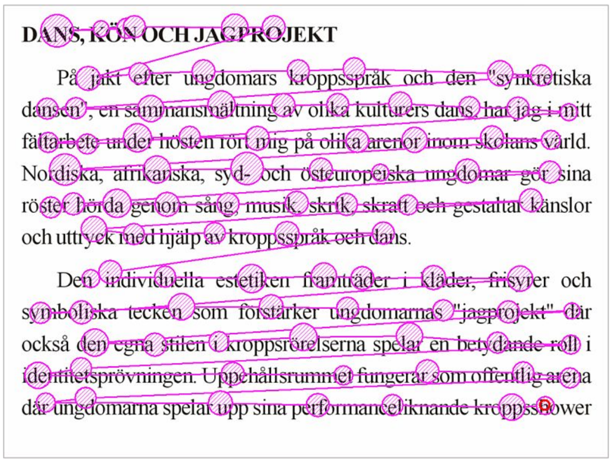
Additionally to the de-identification, some works evaluate the loss of utility. One important goal in regard to human listeners is to achieve a natural-sounding de-identified voice. The naturalness is evaluated by human listeners using the mean opinion score. Another important aspect is the intelligibility of the de-identified speech. Intelligibility can be evaluated via human listeners or machine learning models using the word error rate, phoneme error rate, or short-time objective intelligibility. A common limitation we observed is that most evaluations use the clear data to train the recognition model and then test it against the anonymized data. This approach implicitly assumes that the attacker is not aware of the anonymization and hence does not try to circumvent it.

It's worth noting the VoicePrivacy challenge [358], an initiative to improve the methodology of speaker anonymization. They use EER and the log-likelihood-ratio cast function (Cllr) to evaluate speaker verifiability and word error rate to evaluate speech intelligibility. In a post evaluation, they also retrained their speaker verification systems with anonymized speech data to test against an informed attack. In recent years (since 2020), the VoicePrivacy Challenge framework has become a popular choice for evaluating voice anonymization. The baselines of the challenge have also often been used as the basis for new anonymization techniques.

Qian et al. [299] present a framework to reason about the privacy and utility of voice anonymization techniques. They present the measure of p-leak limit that should give a maximum privacy leakage per speaker for a published dataset. Zhang et al. [402] propose a theoretical framework to quantify the privacy leakage risk and utility loss for speech data publishing.

A.2. Eye Gaze

Eye gaze involves two types of movements: **fixations** and **saccades**. Our eyes alternate between them during visual tasks, such as reading (see Figure A.2). Fixations refer to maintained visual focus on a single stimulus, while saccades are rapid eye movements between fixations to reorient our gaze. Besides, even during fixations, our eyes are not completely still, but constantly producing involuntary micro movements (hundreds per second) known as microsaccades [5].



DANS, KÖN OCH JAGPROJEKT

På jakt efter ungdomars kroppsspråk och den "synkretiska dansen", en sammansmältning av olika kulturers dans, har jag i mitt fältarbete under hösten rört mig på olika arenor inom skolans värld. Nordiska, afrikanska, syd- och östeuropeiska ungdomar gör sina röster hörda genom sång, musik, skrik, skratt och gestaltar känslor och uttryck med hjälp av kroppsspråk och dans.

Den individuella estetiken framträder i kläder, frisyrer och symboliska tecken som förstärker ungdomarnas "jagprojekt" där också den egna stilen i kroppsrörelserna spelar en betydande roll i identitetsprövningen. Uppehållsrummet fungerar som offentlig arena där ungdomarna spelar upp sina performanceliknande kroppsshower

Figure A.2.: Fixation and saccades while reading, from a study of speed reading made by Humanistlaboratoriet, Lund University, in 2005. Source:http://en.wikipedia.org/wiki/File:Rea.

Eye-tracking technologies are becoming increasingly available in the consumer and research market. The most common type of tracking technology works by illuminating the eye with an array of non-visible light sources that generate a corneal reflection. These reflections are sensed and analyzed to extract eye rotation from changes in reflections. There is a wide range of hardware configurations for eye-tracking, including embedded cameras in computers, smartphones and virtual reality headsets, dedicated external hardware, or mobile eye-wear. These sensors allow to extract measurements not only regarding movement data related to fixations and saccades (speed, gaze angle, attention spots, scan path), but also additional features, such as pupil size variations and blink behavior. Combinations of these features provide valuable information to implement eye gaze driven applications.

A.2.0.1. Utility

Eye movements have been studied, analyzed and used for more than a century in different research domains. In the medical field, gaze provides useful information about our cognitive and visual processing [118, 22], which can be used for diagnosing different diseases. In computer science, eye gaze is used as a form of human computer interaction to improve accessibility, user experience, and to adapt system behavior [211, 289, 58]. More recently, security and privacy researchers have focused on analyzing stable unique features of eye movement to build biometric authentication systems [158]. Behavioral eye biometrics have been subject of intense investigation in the last decade, showing EERs as low as 1.8% [82]. Across all these different domains, the utility to be preserved would depend on the underlying application, e.g., accuracy in predicting the next eye movement, in diagnosing a mental disease, in detecting the focus of user attention, or in recognizing a user.

A.2.0.2. Threat Space

Eye movement data is rich in information that can be exploited by malicious entities or curious service providers to uncover user sensitive attributes beyond those disclosed intentionally and required for the purpose of the service or to directly identify a person. Besides the biometric information carried by eye movement data, research has also documented their correlation with multiple disorders and mental conditions, such as Alzheimer's [142], schizophrenia [182, 130], Parkinson [174] bipolar disorder [98], mild cognitive impairment [385] multiple sclerosis [74], Autism [39, 372], or psychosis [86], to name a few. Furthermore, pupil size is known to be an indicator of a person's interest in a scene [123] and a proxy for detecting cognitive load [220, 172]. Other recent works demonstrated that eye data can be used to infer gender and age, or even personality traits [173, 31]. Given the richness of eye data and the increased availability of consumer tracking devices and the advent of eye gaze driven applications, there is a significant and imminent privacy threat potential [7]. The privacy threats of eye-tracking technologies have also been recognized by hardware makers like Apple, which disallow the usage of eye-tracking information for third party applications in their Vision Pro Headset.

The two main threats that endanger eye privacy are re-identification and attributes' inference.

A.2.0.3. Anonymization Techniques

We found multiple recent proposals to protect the privacy of eye movement data, with many of them using noise injection to achieve differential privacy (DP).

Random Perturbation David-John et al. [65] adapt the task-based marginal model for eye gaze, in which for each feature vector dimension a distribution of the values is built to then random sample new synthetic data from these distributions. The identification accuracy of the generated synthetic data is close to chance level.

Noise Injection Steil et al. [340] propose a DP-based technique to protect eye movement data collected while users read different types of documents (comic, newspaper, textbook) in a VR setting. The utility goal is to accurately predict the type of document to provide enhanced features in the reader application. Additionally, the privacy goals are to avoid gender inferences from eye movement data and to protect against re-identification when the attacker has prior knowledge of a data set including the target user's eye data and identity. To achieve these goals, the exponential mechanism [81] is applied to a database of users' eye features by a trusted curator prior to its release. This sanitized database can be then used for training classifiers to provide the enhanced reader functionality. The experiments testing at various noise levels shows that utility with regard to document classification can be partly preserved (~55-70%) while reducing gender accuracy inference to the level of random guesses (~50%).

Based on Steil et al.'s data set, Bozkir et al. [40] evaluate two types of DP-based perturbations, the standard Laplacian perturbation algorithm (LPA) [80] and the Fourier perturbation algorithm (FPA) [303]. They also propose a modification of the FPA algorithm that splits eye data in chunks before adding noise, in order to reduce temporal correlations, which is a source of reduced utility as more noise is required to protect privacy. With this modification, they obtain document type classification results similar to those used by Steil et al. [340] for the case of 50% gender classification, while adding more noise to the data (better privacy guarantee).

Liu et al. [187] present a DP-based solution to anonymize eye tracking data aggregated as a heatmap. A heatmap, or attentional landscape, is a popular method for visualizing eye movement data that represents aggregate fixations [79]. This means that the intensity of every pixel is adjusted relative to the number of fixations over that region. The privacy goal in this case is to protect individual gaze maps while preserving the utility of the aggregated heatmap. Their experiments with random selection and additive noise (Gaussian, Laplacian) show that Gaussian noise is the best option to obtain good privacy guarantees for the individuals' gaze maps without visually distorting the hotspots in the aggregated heatmap, i.e., keeping a certain utility.

David-John et al. [67] worked on protecting eye tracking data recorded in VR/AR headsets. They propose two different interface models for how data can be shared with a third party and propose three anonymization techniques, Gaussian noise injection, temporal down sampling, and spatial down sampling for one of the interface models. The noise injection approach was found to be the most effective as it reduced the identification rate of the subjects the most with high variance values for the Gaussian distribution. Wilson et al. [375] also proposed adding Gaussian noise to eye tracking data, showing similar results.

Hu et al. [136] proposed a local differential private mechanism for generating synthetic eye movement trajectories called Otus. Their technique first separates the field of view into tiles and then constructs a graph that encodes the gaze duration of each tile and the transition probability between the tiles. The graph is then perturbed using the Laplacian mechanism before it is sent to the server. The server then averages all users graphs and uses random walks on the graph to generate new eye movement trajectories.

Li et al. [183] proposed Kalϵido a plugin system that can be used to anonymize eye gaze trajectories with differential privacy guarantees. The authors extend geo-indistinguishability [16] and w-event privacy [159] to take into account the area of interest with radius r a user is looking at. The intuition of their guarantee is that all gaze positions within the area are indistinguishable. They note that they only protect against spatial information and not temporal information. Further, they define an adaptive algorithm to allocate the privacy budget of a user depending on the total privacy budget of each time window. Their results show a reduction of the identification of users to near chance level, however, the utility of the data is also close to chance level.

Coarsening The temporal and spatial down sampling proposed techniques by David-John et al. [67] are both coarsening based techniques. For the temporal down sampling only a very small reduction in the identification accuracy can be recorded while the spatial down sampling has a bigger effect but must be scaled very high to do so. Wilson et al. [375]

proposed a spatial down-sampling approach for the eye gaze angle. They first map the 180° to 2160 points and then coarsen the gaze angle to these points. In their evaluation, the spatial down-sampling seems to be more effective than temporal down-sampling.

Continuous Conversion Wilson et al. [375] propose smoothing the eye gaze using a sliding window approach. They show that using a large enough window reduces the identification rate.

David-John et al. [65] applied k-anonymity to eye movements by grouping the trajectories of users and then averaging them. They were able to show that even with small numbers of k the identification accuracy drops significantly. Due to them processing the feature vectors of each task separately their reported high utility is questionable. In a follow up paper, David-John et al. [66] propose two synthetic data generation approaches for eye gaze. Their k-same synth approach applies k-anonymity to the fitted parameters of a Gaussian mixture model before using it to generate fixations and saccades. Their event-synth-PD approach uses a conditional variational autoencoder to generate new data with given characteristics. They show that their event-synth-PD approach achieves plausible deniability. They compare both methods to Kalϵido and achieve comparable results for privacy and utility.

Fuhl et al. [95] perform eye gaze anonymization by using an auto encoder in combination with reinforcement learning. The auto encoder is trained on the eye gaze trajectories to learn a latent representation of the data. Then a manipulation agent modifies the latent vector of the trajectories to prevent for example gender classification. After the decoding of the latent vector, a classifier tests how good the manipulation was and its result is used as the loss for the training of the manipulation agent.

A.2.0.4. Evaluation

The proposals by Steil et al. [340] and Bozkir et al. [40], measure the quality of their anonymization techniques for attribute inference protection using the classification accuracy metric for the main task and the attribute inference task. For the re-identification protection case, it is assumed that the attacker has previous knowledge of a database of users' eye data and their identities. To simulate this knowledge, they train the classifiers on the clean data and test them on the anonymized data, using also the accuracy metric to report privacy protection. Besides, these works also report the so called privacy loss parameter (or ϵ) from DP theory, which quantifies the maximum difference between the data points of two individuals in the data set. Furthermore, Bozkir et al. use the inverse of the normalized mean square error (NMSE) between the actual eye feature values and the perturbed ones as a utility metric. However, the interpretation and implications of these privacy loss and utility metrics are not developed.

Liu et al. [187] analyzed the privacy-utility trade-off of anonymized heatmaps using the correlation coefficient (CC) and mean square error (MSE) of noisy heatmaps under different privacy levels (different values of ϵ). The CC and MSE give an idea of the similarity between the original and the anonymized heatmaps and the ϵ provides information about the privacy guarantee (the smaller, the better privacy). These metrics are accompanied by the visual representation of the noisy heatmap, in order to aid the relevant stakeholders in deciding what level of noise is acceptable for a given application.

Regarding datasets, the largest dataset available is GazeBaseVR [192], which captured 407 participants performing 5 tasks with up to 6 sessions. As recording device they used a VR headset. Steil et al. [340] collect data from 20 participants (10 male, 10 female, aged 21-45) while reading documents using a VR headset. Each recording is divided into three sessions (reading a comic, newspaper, or textbook), lasting 30 minutes in total. They extract 52 eye movement features related to fixations, saccades, blinks, and pupil diameter. The dataset has been publicly released [1] by the authors and Bozkir et al. [40] use it as the basis to evaluate their proposal.

The Ehtask [137] dataset contains the recordings of 30 people performing 4 different eye gaze tasks using a VR headset. Another VR headset dataset is DGaze [78], which captures 43 people in 5 different scenes. In the heatmaps anonymization study, Liu et al. use a synthetic simulated dataset to illustrate their privacy analysis. Besides the technical privacy analysis, Steil et al. [340] is one of the few works considering user privacy concerns regarding behavioral data collection. They conduct a large scale user survey (with N=164 participants) to explore with whom, for which services, and to what extent users are willing to share their gaze data. Their report shows that people are uncomfortable with inferences (gender, race, sexual orientation) and would object to sharing their data if these attributes can be leaked. The results also show that people generally agree to share their eye tracking data with a governmental health agency or for research purposes, but would object to doing so if the data owners are companies. These insights are a first step towards understanding user privacy awareness and privacy needs, but more work is required in this field to guide the design of user-centered privacy protective techniques for behavioral data.

A.3. Brain Activity

Brainwaves are patterns of measurable electrical impulses emitted as a result of the interaction of billions of neurons inside the human brain. Since the first human electroencephalogram was recorded in 1924 [111], both the hardware devices to measure brain activity and the analysis techniques to process these signals have significantly improved. Current technologies to measure brainwaves can be classified as invasive and noninvasive methods. Invasive methods record signals within the cortex by directly implanting electrodes near the surface of the brain [156]. These methods are far too risky for usage under noncritical circumstances and are only used in clinical applications. Instead, non-invasive methods are most frequently used and applicable to many areas other than the medical realm, such as brain-controlled interfaces. The most portable and commonly used of these techniques is electroencephalograpy (EEG), which records electrical activity through sensors placed on the scalp surface.

An EEG signal is a combination of different brainwaves occurring at different frequencies. Every type of wave carries different kinds of information, which can be used to gain insights about the current state of the brain [11]. Researchers have tried to identify certain mental states associated to each brainwave. Table A.1 presents a summary of the most important wave types, their respecttive frequencies, their originating location in the brain, and their associated mental state.

Brain-computer interface (BCI) technologies mostly work on continuous EEG data recordings, i.e., time series data. But there are also many applications based on the extraction

[1] https://www.mpi-inf.mpg.de/departments/computer-vision-and-machine-learning/research/visual-privacy/privacy-aware-eye-tracking-using-differential-privacy

of time-locked brain variations that appear in reaction to external stimuli. These variations, called event related potentials (ERPs), are widely used to detect neurological diseases. In both cases, either using ERPs or a longer EEG series, features are computed for the brainwave data-driven application built on top. These features can belong to the time and/or frequency domain and to one or multiple channels. Examples of commonly used features include Autoregressive coefficients, Fourier and Wavelet transforms.

Table A.1.: Overview of EEG brainwaves - based on [11] and [3].

Wave Type	Freq. (Hz)	Originating Location	Mental State
Gamma γ	30-100	Somatosensory cortex	Active information processing, strong response to visual stimuli [3]
Beta β	13-30	Both hemispheres, frontal lobe	Increased alertness, anxious thinking, focused attention
Alpha α	8-13	Posterior regions, both hemispheres; High amplitude waves	Resting, eyes closed, no attention [161]; Most dominant rhythm
Theta θ	4-8	No special location	Idling, dreaming, imagining, quiet focus, memory retrieval
Delta δ	0.5-4	Frontal regions; High amplitude waves	Dreamless and deep sleep, unconsciousness

A.3.0.1. Utility

The utility that should be preserved when processing brainwave data is highly dependent on the application. For clinical applications, for example, the raw information could be needed for a proper diagnosis or a safe brain controlled prosthesis. In these cases, regulations like the HIPAA Privacy Rule [126] are usually in place to protect personal identifiable information. When moving to other less regulated fields of application, the need for full raw EEG data is not necessarily justified. The most prominent EEG applications include user authentication, personalization of gaming experiences, and brain controlled-interfaces. In these cases, the utility to be preserved should be enough to provide a useful application, i.e., recognize the user, and offer personalized options and responsive interfaces all with a tolerable error that does not hamper the security and usability of the service.

A.3.0.2. Threat space

Brain activity is rich in information. It can be used to uniquely identify individuals given their unique characteristics and, in fact, several biometric systems based on brainwaves have been proposed [108]. Besides, the acquisition of EEG signals raises privacy issues because brainwaves correlate, among others, with our mental states, cognitive abilities, and medical conditions [345]. Martinovic et al. [216] demonstrated that by manipulating the images presented to the users, their EEG signals could reveal private information, e.g., bank cards, PIN numbers, area of living, or if the user knew a particular person.

A.3.0.3. Anonymization Techniques

We found that a large number of anonymization rely on machine learning methods to perform the anonymization of the data, with approaches like Generative Adversarial Networks

(GANs) and adversarial perturbation scheme dominating the field. With the availability of EEG datasets the anonymization of brain activity data is gaining some traction.

Feature removal Matovu et al. [218] explore how to reduce the leakage of private information from EEG user authentication templates. They assume an insider type of attacker, such as an unscrupulous database administrator, who misuses their privilege to maliciously exploit the templates. The attacker wants to infer, specifically, if the user associated with a template is an alcoholic. Their envisioned anonymization technique aims at concealing the alcoholism information while still providing good authentication accuracy. It is, therefore, an attribute protection mechanism. Conceptually, it lies on the hypothesis that different template designs (features, channels, frequencies) will have an impact on the amount of non-authentication information (emotions, health conditions) that can be inferred. The authors demonstrate this hypothesis by choosing two different templates and calculating the predictive capability to authenticate users and determine their alcohol consumption behavior.

Continuous Conversion In the same direction of feature selection, Yao et al. [391] propose the usage of Generative Adversarial Networks (GANs) [103] to filter sensitive information out of EEG data. Their goal is to reduce the possibility of inferring alcoholism while keeping the brain activity recordings useful to detect mental tasks, specifically to predict which visual stimulus the user is looking at. The GAN-based proposed filter involves deep neural networks that perform domain transformation, that is, translating EEGs from a source domain distribution X with both desired and privacy-related features to a target domain distribution Y with desired features only. Their results after applying the filtering technique show a significant reduction in the percentage of EEG sequences from alcoholic users that can be classified as such (from 90.6% to 0.6%). At the same time, the mental task classification accuracy does not drop significantly (4.2% less). However, the original mental task classifier accuracy was not strong before filtering the privacy-sensitive features and it remains to be studied if this technique would work in other classification scenarios.

Pascual et al. [273] use a GAN to generate synthetic EEG data to train an epilepsy monitoring system as sharing large amounts of medical EEG is a privacy problem. The authors focus on inter-ictal EEG signals (signals between two seizures) as these are easier to record than the actual seizures. As generator a convolutional auto encoder is used but instead of decoding an inter-ictal the latent code is translated into an ictal sample. The discriminator then compares the synthetic ictal to a real one. Their results show that the synthetic data reaches identification rates which are close to chance level, even when only two patients are in the test set. However, this is only a pseudonymization of the patients as all synthetic ictal values generated for a specific patient can still be linked to each other.

Bethge et al. [32] proposed privacy encoders to remove the sensitive information from each of the brain activity data streams before they are used in a classification task. For each data set a convolutional neural network is trained as encoder using the maximum mean discrepancy (MMD) between the different encoded data sets as loss function. This way the encoders should learn a domain-invariant representation of the data. They test their approach on four data sets finding that the classification from which data set a sample originated drops from 99% to 52%, while the emotion classification is only reduced from 51% to 49%. It remains an open question how well the identity of a subject would be preserved by this approach. A similar approach is being proposed by Meng et al. [229], instead of using a neural network for the transformation, they learn a perturbation vector that is added to the EEG signal. The perturbation is learned via an adversarial scheme using an action classifier to establish the utility and a biometric recognition system for the privacy. In addition to learning one perturbation vector per EEG sample they show that it is possible to learn such a perturbation vector for each user, allowing for fast anonymization of unseen EEG samples of known users. Another adversarial approach is being proposed by Singh et al. [334]. The main difference from the previous approaches is that an autoencoder is used for the transformation.

Continuous Conversion + Noise injection Debie et al. [68] also use a GAN to generate new synthetic data from the original one. They differ from Yao et al. and Pascual et al. in that they use differentially private stochastic gradient descent on the discriminator of the network. This method reduces the influence of each individual to the computation of the gradients. They evaluated their GAN on the Graz data set A with EEG data from 9 subjects. Their results show that the utility of the synthetic data is well preserved, however, no additional privacy evaluation was performed.

A.3.0.4. Evaluation

The reviewed works, similar to the proposals for anonymizing gait, evaluate the quality of inference protection by comparing the prediction accuracy for the protected attribute before and after modifying the EEG data. The metrics used for this analysis are typical machine learning metrics, including accuracy, false positive rates, and false negative rates. Similarly, the loss of utility is evaluated by measuring the reduction in classification accuracy when using the original and anonymized EEG data.

For their evaluations, the works use a variety of different EEG datasets. The largest dataset is the Temple University Hospital EEG data corpus [263] which contains 579 subjects, followed by the BCI2000 dataset [321] with 106 subjects. Specifically recorded for authentication was the dataset of Arias et al. [19] which recorded 56 people. A special dataset is the SUNY medical dataset with EEG data of 25 alcoholic subjects and 25 control subjects while looking at visual stimuli [258, 157]. Further, there exist a couple of smaller datasets [341, 373, 129].

A.4. Heartbeat

An electrocardiogram (ECG) is a graph of voltage over time that captures the electrical activities of cardiac muscle depolarization followed by repolarization during each heartbeat. Shown in Figure A.3, the ECG graph of a normal beat is composed of a sequence of waves: a P-wave reflecting the atrial depolarization process, a QRS complex representing the ventricular depolarization process, and a T-wave denoting the ventricular repolarization. Other portions of the ECG signal encompass the PR, ST, and QT intervals [407].

Like other biometric systems applied to identification tasks, ECGs are typically converted into abstract, compressed representations, typically referred to as biometric templates, before the task is conducted. Biometric-template methods can be classified depending on the exploited features of the ECG data. The most popular ones are fiducial-based, non-fiducial-based and hybrid methods [264]. On the one hand, fiducial-based techniques utilize characteristic points on the ECG signal to extract temporal, amplitude, envelope, slope and area features. Characteristic points are the locations that correspond to the

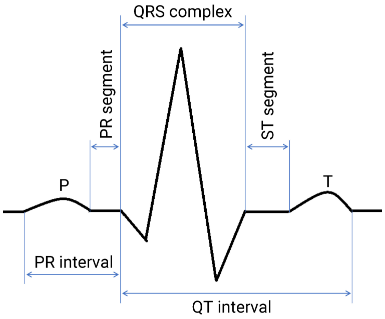


Figure A.3.: Waveform of an ECG signal with normal cardiac cycle. Source: https://www.nottingham.ac.uk/nursing/practice/resources/cardiology/function/normal_duration.php.

peaks and boundaries of the P, QRS and T-waves of the ECG signal. On the other hand, the non-fiducial-based methods do not rely on the ECG characteristic points, and examples include autocorrelation coefficients, Fourier and wavelet transforms. Hybrid methods combine both fiducial-based and non-fiducial-based features.

A.4.0.1. Utility

ECG data find application in healthcare and biometrics systems, the latter being intended for identification and authentication [363]. In healthcare, ECGs are utilized for diagnosis of heart diseases [189]. Typically, there is a stand-alone service or a complete e-health system where the service provider, in addition to offering a repository of personal medical data, may allow to remotely process such data. In any case, the aim is to provide real-time feedback to patients and hospitals, either as a warning of impending medical emergency or as a monitoring aid during physical exercises.

A.4.0.2. Threat Space

Regardless of the application (i.e., identification, authentication or healthcare), ECGs are health data and, as such, are considered sensitive by data-protection regulations and need to be protected. Consider the case, for example, of a user who might see their insurance premium increased or suffer discrimination during a job application due to a medical condition inferred from their ECGs.

Although it is well known that ECG data may help diagnose a patient's physiological or pathological condition, other probably lesser-known inferences include cocaine use [135] and stress [284], which may be sensitive to the patient and obviously should be kept private. The fact that the very same time series data allows drawing both desirable inferences (i.e., for healthcare) and sensitive inferences (that need to be protected) poses a dilemma of great practical relevance.

A.4.0.3. Anonymization Techniques

Next we survey the most relevant privacy-protection techniques for ECG data.

Another approach based on compressive sensing (CS) [45] is proposed by Djelouat et al in [77]. CS is a signal processing technique that combines both sampling and compression through random projections. Building on this technique, the authors propose compressing the ECG signal by sampling it at the time of sensing. This reduces the need to even store the sensitive ECG data at the wearable device, thereby providing protection against that entity. The theoretical properties of this compression technique ensure that, under certain assumptions on the random projection, a good reconstruction of the original ECG signal can be obtained at the provider side.

Feature Removal Kalai et al. [398] present a template protection scheme for ECG data. In a first phase, the authors propose computing the discrete cosine transform (DCT) of the ECG signal's autocorrelation coefficients, and then removing those DCT coefficients with the lowest energy. The remaining DCT coefficients constitute the biometric template. In a second phase, two keys are obtained from the template. One is transmitted to the target application the user wishes to authenticate. The other functions as a private key, which is derived from the complete DCT already stored in the server. A similar approach is presented by Zaghouani et al. [399] that uses a quantization step once the DCT-template is obtained. This latter approach is evaluated on the PTB dataset but no experimental comparison is conducted between the two proposed solutions.

Another similar proposal is made by Mahmoud et al. [208], which decomposes the ECG signal into its wavelet transform, eliminates the low-frequency coefficients and reconstructs the ECG signal for release. At the provider side, only authorized personnel with access to a secret key (derived from the wavelet-transform template) is able to reconstruct the original ECG from the released, protected signal. To which extent these released data may safeguard patients' privacy is evaluated through the percentage root mean square difference (PRD), a simple and widely used distortion measure in ECG signal processing applications [215] that quantifies the difference between the original ECG and its protected version.

Continuous Conversion Bennis et al. [30] proposed a simple k-anonymity scheme for ECG data. In their first step they transform the signal into the frequency domain. Next they pick the k closest neighbours of the signal and then aggregate those into a new signal before transforming it back into the time domain.

Piacentino et al. [282] used a GAN to generate synthetic ECG data by first normalizing the data and then arranging it into a matrix. For the arranging of the data multiple proposals are made sorting the data values by their type. No evaluation of the privacy of the synthetic data was performed. Jafarlou et al. [146] also propose to use a GAN to generate anonymized ECG data samples. Their approach differs from Piacentino et al. in that they use the original ECG sequence as input to the GAN and use the identification accuracy as part of the training loss for the GAN. Their evaluation shows lower identification accuracies while still allowing arrhythmia detection. Nolin-Lapalme et al. [262] also use a GAN for the ECG anonymization, but they aim at generating sex neutral ECG samples and use the sex classification as part of the GAN loss.

Random Perturbation + Noise Injection Although encryption based on the idea of CS can achieve a computational notion of secrecy through the random projection step, it has been shown this technique is vulnerable from an information-theoretic perspective [300]. To address this problem, Chou et al. [55] propose using principal component analysis and SVD on a CS scheme, where the ECG data is encrypted at the wearable sensor by adding signal-dependent noise. They measure privacy as the mutual information between the original ECG signal and its encrypted version, and show that high classification accuracy can be achieved while providing privacy beyond computational secrecy.

Discrete Conversion + Noise Injection Unlike the works surveyed previously, the goal of Zare-Mirakabad et al. [400] is to publish suitable representations of ECG data with certain privacy guarantees. To do this, Zare-Mirakabad et al. propose converting ECG time series into symbolic representations over time. They use the popular Symbolic Aggregate approXimation (SAX) to replace continuous numerical values with strings of symbols (see Figure A.4). With this new symbol representation, the proposed anonymization technique first builds an n-gram model from the complete time-series string, and then ensures that each n-gram has a minimum frequency of occurrence, similar to the k-anonymity criterion. To ensure this version of k-anonymity is satisfied over the string of symbols, the authors contemplate adding fake n-grams to the original string. Experimental results on the Eamonn Discord Dataset show that (a measure of) information loss is hardly affected for values of k up to 20.

Continuous Conversion + Random Perturbation Chen et al. [49] and subsequent work by Wu et al. [378], address the problem of making ECG-based biometric templates revocable, exactly as keys or passwords, a property they consider indispensable in order for ECGs to be used in practice. To enable template revocability, the common practice is to associate distinct templates with the same biometrics by perturbing them in a different manner. To protect user privacy, however, this process needs to ensure the recovery of the original biometric from its template is either infeasible or computationally hard.

Essentially, cancelable templates are obtained as random projections of a user's ECG data block. Unlike common approaches, however, Wu et al. put no restrictions on the generator matrix. Accordingly, the idea is that each realization of this matrix allows cancelling their corresponding templates. Reidentification is then conducted with the multiple-signal classification algorithm [33], reporting rates of over 95% in the Physikalisch Technische Bundesanstalt Database.

A distinct approach by Hong et al. [131], proposes a template-free identification system to prevent any privacy issue from compromised or stolen templates. The system converts ECG-data into images through various spatial and temporal correlations methods and uses deep-learning techniques to train a classifier. The authors conduct experiments on the Pysikalisch-Technische Bundesanstalt database and report identification rates of over 90% with sampling rates of 1 000 Hz.

Continuous Conversion + Noise Injection Sufi et al. [344] propose building templates of the waves P, QRS and T through cross-correlations of the ECG signal. Each of those templates are then obfuscated in a concatenated fashion with additive noise generated synthetically, so that the obfuscation of a wave serves as input to obfuscate the next wave. The upshot are noisy forms of the three waves and noisy templates thereof. All this information constitutes the key available to authorized personnel, who will be able to reconstruct the original ECG from the noisy version (which is shared or made publicly available by the patient or user themselves). Unauthorized personnel, per contra, will only have access to the noisy ECG signal, which, according to the authors, may prevent identity and attribute disclosure.

Huang et al. [139] propose an authentication system that protects the privacy of ECG templates in a database with differential privacy. The authors assume the interactive setting of this privacy notion, where an analyst queries the database to obtain ECG data. Specifically, the analyst is supposed to ask for the coefficients of a Legendre polynomial, that the anonymization system utilizes to fit and compress the ECG signal. Laplace noise

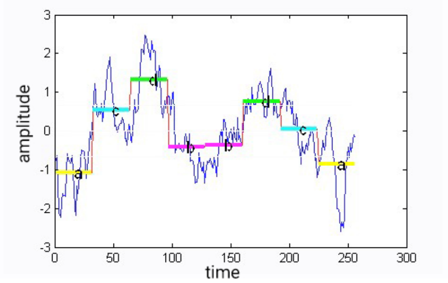


Figure A.4.: A time series is converted into the string "acdbbdca". Source: https://cs.gmu.edu/~jessica/sax.htm.

is calibrated to the sensitivity of those coefficients and added to them, and the noisy response is returned to the analyst. The ε parameter of DP therefore regulates the trade-off between user privacy and authentication accuracy, the latter aspect depending on two sources of error: the polynomial fitting approximation and the injected noise. The authors evaluate the system in the MIT-BIH ECG and MIT-BIH Noise Strees databases, reporting decent authentication accuracy. However, they appear to misunderstand how the sensitivity of the coefficients is computed and therefore their results seem to have been obtained incorrectly.

Saleheen et al. [318] investigates if sensitive inferences from segments of time series data can be drawn by a dynamic Bayesian network adversary. The adversary is assumed to estimate a range of behavioral states about the user, including, for example, whether or not they are in a conversation, running, smoking and stress, at the time the data is gathered. When the adversary is likely to infer sensitive aspects of a user, the corresponding segments of data are substituted for most-plausible, non-sensitive data. To estimate the privacy provided by these substitutions of data, the authors propose a variation of the differential-privacy notion that bounds the information leaked resulting from the substitutions. In other words, the proposed metric ensures that the information leaked about a sensitive inference from a substituted segment is always bounded. Utility loss is, on the other hand, computed as the absolute difference between the probability of inference about each non-sensitive behavioral state from actual data, and the same probability from released data. Although experimental results show relatively small values of utility loss for $\varepsilon \in [0.05, 0.65]$, the proposed solution has two main limitations: first, protection is provided only for dynamic Bayesian network adversaries; and secondly, it assumes all time-series data are available beforehand, which precludes its application in real-time scenarios.

A.4.0.4. Evaluation

The reviewed techniques measure how service functionality is degraded due to anonymization with common machine learning metrics like precision, recall and accuracy, and less frequently with the DTW and PRD quantities, which assess the similarity between original and protected time series. As for privacy, the level of protection is assessed through a variety notions and measures, including the accuracy of a membership inference attack, the ε parameter of differential privacy, the mutual information between the original ECG signal and its encrypted version, the probability of correct inferences on sensitive attributes with and without protection, and through a notion similar to k-anonymity. A common dataset used is the MIT-BIH arrhythmia database [243] which contains the ECG samples of 47 people.

A.5. Discussion

Table A.2.: An overview of all found methods classified by trait and method. Papers that propose multiple methods can appear in multiple rows. Papers that combine multiple methods are marked the following: * plus noise injection, † plus random perturbation, ‡ plus discrete conversion.

Trait Method	Voice	Gait	Hand motion	Eye-Gaze	Heartbeat	Brain activity
random perturbation	[270] [248] [329]	[128]	[209] [105] [210] [367]	[65]	[55]*	
noise injection	[347] [120] [113] [265] [364] [204] [114]	[352] [353] [219] [115] [231]	[237] [325]	[340] [187] [183] [67] [136] [375]		
coarsening		[252]	[210] [367]	[67] [375]		
feature removal	[272] [271] [379] [404] [255] [57] [76] [256] [8]	[154] [99] [115] [69] [313]			[398] [399] [208]	[218] [391]
discrete conversion	[274] [290] [291] [34]		[317] [180] [238] [367] [92] [249]		[400]*	
continuous conversion	[150][285][340] [155][4][295][23] [89][160][90][1] [205][197][338][109] [12][394][293][10] [275][221][52][266] [235][125][386][389] [234][257][47][233] [397][203][388][236] [235][280][390][48] [72][56][381][304] [222][100][294]† [333]†[326]*[46]† [167]*[168]*[297]* [298]*[338]*	[9] [144] [350] [112] [244] [251] [127]‡	[209] [213] [320] [380] [87]	[65] [95] [375] [66]	[30][282] [146] [262] [49]† [378]† [131]† [344]* [139]* [318]*	[273] [32] [229] [334] [68]*

Table A.3.: An overview over which privacy goals the different techniques try to achieve.

Trait Priv. Goal	Voice	Gait	Hand motion	Eye-Gaze	Heartbeat	Brain activity
Attribute	[248][113][204][274][290][291][34][295][90][12][46][56][125]	[128][99][115]	[209][105][210][367][317][238][209]	[340][40][95]	[398][399][208][49][378][131][344][139][318]	[218][391][32][68]
Identity	[270][271][120][113][204][265][364][114][272][379][404][255][256][57][150][285][340][155][4][295][23][89][221][293][160][1][197][205][12][338][10][394][275][109][167][168][297][298][329][8][52][266][235][125][386][389][234][257][47][233][397][203][388][236][235][280][390][48][72][381][304][222][100][294][333][326]	[219][352][353][354][154][99][9][144][350][127][115][231][252][69][313][112][244][251]	[237][325][180][92][249][213][320][380][87]	[65][340][40][187][67][136][183][65][95][375][375][375][66]	[30][282][55] [400][146] [262]	[273][68][229][334]

B. Motion Data Collection

B.1. CeTI-Locomotion

B.1.1. Usage Notes

The motion data and additional metadata are stored as tab-separated values (.tsv) and JSON (.json) files, and are as such compatible with a wide range of software applications and programming languages, thus promoting interoperability and seamless integration in different environments, as well as data processing efficiency. We recommend to use the processed data in *derivates/cut_segments* for analysis and classification, as the single repetitions already contain a lot of information. For the comparison of the classification of the identity or the actions performed, the code provided for the verification experiments can be used as a simple baseline. For the BIDS standard [104] there are specialized libraries, such as PyBIDS [392, 393], for importing and processing the data.

B.1.2. Code availability

The custom code used for processing and technical validation are available along with the associated dataset [116]. The required libraries to execute the custom scripts have been included in the files *requirements.txt* and *requirements.yaml*. These files allow for the installation of the necessary libraries either directly via The Python Package Index (PyPI)[1] or via the Anaconda[2] software distribution (2020.11). The script *preprocess_data.py* encompasses all the data processing steps that were performed subsequent to the export from Rokoko studio. To facilitate the replication of the technical validation, the script *verification_experiments.py* has been provided. This script allows for the execution of the technical validation on the processed data. Furthermore, the script *render_sequence.py* allows for the rendering the position data from *_motion.tsv files for visual analysis and verification. For additional details on the usage and execution of the custom code, please refer to the *README* file, which provides comprehensive instructions and guidelines.

B.2. FacialMotionID

B.2.1. Additional Tables

B.2.2. Text-Level Segmentation

We further segmented the verbal tasks into words (nursery rhymes only) and phonemes. To accomplish this, we aligned the speech recordings collected during task execution with the transcript of the performed task. We used a force alignment model to automatically

[1] https://pypi.org

[2] https://www.anaconda.com

Table B.1.: Overview of the different tasks which the participants performed in the study. v: verbal, nv: non-verbal

ID	Type	Task	Repetitions
0	v	sixpence (word)	4 / 5
1	v	dinosaurs (word)	4 / 5
2	v	muffin (word)	4 / 5
3	v	Sing a Song of Sixpence (rnhyme)	4 / 5
4	v	Dinosaurs (nrhyme)	4 / 5
5	v	The Muffin Man (nrhyme)	4 / 5
6	nv	happiness	4 / 5
7	nv	anger	4 / 5
8	nv	fear	4 / 5

Table B.2.: Mapping from the device-dependent motion data attributes to the unified data format. For n-to-1 mappings from the devices to the unified format we use the mean of the directions.

Type	Unified	Vive	Pico	Meta	Direction (*)
Facial	CheekPuff	Cheek_Puff_*	CheekPuff	CheekPuff*	left/right
	EyeClosed*	Eye_*_Blink	EyeBlink_*	EyesClosed*	left/right
	EyeLook*	Eye_*_*	EyeLook*	EyesLook*	left/right, down/up, in/out
	Jaw	Jaw_*	Jaw*	Jaw*	forward/thrust, left/right, open/drop
	LidTightener*	Eye_*_Squeeze	EyeSquint_*	LidTightener*	left/right
	UpperLidRaiser*	Eye_*_Wide	EyeWide_*	UpperLidRaiser*	left/right
	LipCornerDepressor*	Mouth_Sad_*	MouthFrown_*	LipCornerDepressor*	left/right
	LowerLipDepressor*	Mouth_Lower_Down*	MouthLowerDown_*	LowerLipDepressor*	left/right
	UpperLipRaiser*	Mouth_Upper_Up*	MouthUpperUp_*	UpperLipRaiser*	left/right
	LipCornerPuller*	Mouth_Smile_*	MouthSmile_*	LipCornerPuller*	left/right
	LipPucker*	Mouth_Pout	MouthPucker	LipPucker*	left/right
	LipSuckB	Mouth_Lower_Inside	MouthRollLower	LipSuck*B	left/right
	LipSuckT	Mouth_Upper_Inside	MouthRollUpper	LipSuck*T	left/right
	Mouth*	Mouth_*_*	Mouth*	Mouth*	lower/upper
	TongueOut	Tongue_LongStep*	Mouth*	Mouth*	lower/upper
Eye	LookDirection*	Gaze_Direction_*	LookDirection*	LookDirection*	X,Y,Z; left, right
	Position*	Gaze_Origin_MM_*	Position*	Position*	X,Y,Z; left, right
Head	DevicePosition*	DevicePosition*	DevicePosition*	DevicePosition*	X,Y,Z; left, right
	DeviceRotation*	DeviceRotation*	DeviceRotation*	DeviceRotation*	X,Y,Z,W; left, right

Table B.3.: Overview of the optimized parameters

Parameter	Range	Note
Layer Size	10-256	Only Simple & LSTM
Hidden Layers	0-2	Only Simple & LSTM
Learning Rate Step Size	10-100	All
Learning Rate Alpha	0.01-1	All
Optimizer Learning Rate	0.0001-0.1	All
Weight Decay	0.00001-0.01	All

perform this process on all verbal tasks and obtain the offset times for each word and phoneme uttered by the participant.

Due to synchronization problems between the audio recording and the recorded motion data, we first create a transcript of the entire recording by using WhisperX [24], an *Automatic Speech Recognition* (ASR) model, instead of aligning the recordings exclusively with

Table B.4.: Device type recognition accuracy using a participant-wise split for all headsets

Model / Data Type	Simple	LSTM	EKYT	Chance
Facial	1.0	1.0	1.0	0.48
Eye	1.0	1.0	1.0	0.48
Head	1.0	1.0	1.0	0.48

Table B.5.: Verbal task recognition accuracy using a participant-wise split for all headsets

Model / Data Type	Simple	LSTM	EKYT	Chance
Facial	0.78	0.89	0.96	0.17
Eye	0.56	0.6	0.68	0.17
Head	0.17	0.17	0.47	0.17

the text of the verbal tasks. Another benefit of this approach is that we can also account for unforeseeable words that were possibly uttered at the beginning of the recording, and for which we did not have a transcript before. Then, we locate the verbal tasks in the transcript and correct any errors using the text of the specific task.

These transcriptions were then used as input for the *Montreal Forced Aligner* (MFA) [224], along with the full recordings. By being given the full transcriptions, the model accurately aligned them to the audio recordings and returned the offsets of when each word and phoneme was uttered.

As a last step, we had to convert the alignment offsets in the audio recordings to the actual timestamp ranges in the motion data files. To do this effectively, we interpolated the start and stop timestamps of the text tasks in the data with the start and stop alignment offsets of the same text tasks obtained from MFA. As a result, we could segment the text task data into word and phoneme segments.

B.2.3. Additional Experiments

In Experiment **E5**, we investigate if the MR headset type can be inferred from the data collected. All headset data has the same format due to the unified data format, however, we expect that it is easy to infer which headset is being used due to device specific quirks. Then, we look at the inference of sensitive attributes about the user of the MR headset in Experiment **E6**. Here, we seek to infer the sex, English level, and personality trait of the user.

Lastly, we perform two experiments to better understand the identification from facial motion data. In Experiment **E7**, we perform the identification only on the verbal tasks or only on the non-verbal tasks to see which task type works better for identification. And in Experiment **E8**, we test how good we can identify individuals when we combine the facial motion data with the eye gaze and head motion data.

Results Since we have multiple devices, we tested whether we could identify which headset was used to record the data for Experiment **E5** (see Table B.4). Unsurprisingly, we can achieve 100% recognition accuracy for all data types.

Table B.6.: English level recognition accuracy using a participant-wise split for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Facial	0.72	0.73	0.69	0.73
Eye	0.73	0.73	0.71	0.73
Head	0.73	0.73	0.68	0.73

Table B.7.: Personality recognition accuracy using a participant-wise split for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Facial	0.43	0.49	0.41	0.49
Eye	0.44	0.47	0.41	0.49
Head	0.49	0.49	0.38	0.49

In addition to recognizing emotions, we test whether the text task can be identified from the recorded data (see Table B.5). The best recognition accuracy of 96% is again achieved using facial motion data.

We examine the results of the attribute inferences tested in Experiment **E6**. Table B.6 shows the results for English level recognition, and Table B.7 shows the results for classifying whether someone is an ambivert, extrovert, or introvert. For both attributes, the results are close to the level of chance, so we do not believe they can be inferred from the data. For sex recognition, shown in Table B.8, there appears to be some information which can be extracted. Since the EYKT model achieved significantly less than chance level, and with only two classes (everyone identified as either male or female) in the dataset, we can simply invert the labeling.

To better understand which task type is better for identifying individuals, we ran identification Experiment **E7** on only the verbal and non-verbal tasks. See Tables B.9 and B.10 for a comparison. Our results show that verbal tasks perform better than non-verbal tasks. However, it also does not appear that a reliable sex recognition can be implemented with facial motion data for now.

Lastly, we further investigated the identification potential of the data we collected. In Experiment **E8** (see Table B.11), we tested the identification accuracy using all data types simultaneously. We found that combining the three data types increased identification accuracy to 99%, thereby outperforming the best all-headset result from Experiment E1 (see Table 5.2).

Table B.8.: Sex recognition accuracy using a participant-wise split for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Facial	0.65	0.81	0.67	0.81
Eye	0.72	0.81	0.58	0.81
Head	0.73	0.81	0.56	0.81

Table B.9.: Identification accuracy using a random split for only verbal tasks for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Face	0.72	0.65	0.93	0.01
Eye	0.52	0.56	0.83	0.01
Head	0.79	0.77	0.95	0.01

Table B.10.: Identification accuracy using a random split for only non-verbal tasks for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Facial	0.63	0.47	0.8	0.01
Eye	0.38	0.33	0.75	0.01
Head	0.64	0.41	0.89	0.01

Table B.11.: Identification accuracy using a random split for all headsets

Model Data Type	Simple	LSTM	EKYT	Chance
Facial + Eye + Head	0.89	0.83	0.99	0.01