Aalto-yliopisto
Aalto-universitetet
Aalto University

ELEC-E7261
Ambient Intelligence

Data Gathering and Curation:
Body positions and movement
Project Report

Saeideh Mansouri
saeideh.mansouri@aalto.fi
Anh Pham
anh.q.pham@aalto.fi
Aitor Urruticoechea Puig
aitor.urruticoecheapuig@aalto.fi
January-March 2024

Abstract

This report unveils a detailed full-body gesture database. It includes ECG, EMG, and accelerometer data associated with a variety of activities such as writing, typing, phone scrolling, walking, jogging, scouting, and 12 common gestures described in the MSRC-12 dataset [1, 2]. The data was gathered from 20 participants. The data collection process involved the use of shimmer sensors [3], Trigno Quattro [4], and two smartphones using Phyphox [5] application, with an emphasis on capturing full-body movements. Five shimmer sensors were tactically positioned at five distinct points on the participants' upper bodies. Moreover, the Trigno Quattro Sensor was employed on the participants' dominant arms to facilitate flexible muscle tracking, particularly during writing activities. This report provides an overview of the gesture capture system, the organization of the database, its potential uses, and a guided navigation through the database; which is publicly available in GitHub after anonymisation [6].

Contents

List of Figures

List of Tables

Acronyms

1 Introduction

The rapidly developing field of Ambient Intelligence provides innovative insights into human behaviour and interactions with their surroundings. Developments influence it in sensor technology and data analytics. In the pursuit of advancing knowledge within the realm of Ambient Intelligence, the group presenting this report, comprised of three Aalto masters students, embarked on a project addressing the intricacies of sensor detection within various actions. The goal of this study is to explore the complexity involved in capturing raw and processed data that are necessary for further processing steps. An essential component of this project lies in learning how to fully comprehend and utilize a variety of sensors, including Electrocardiogram (ECG) - Shimmer sensors, Electromyography (EMG) sensors, accelerometers, and associated data capture software such as Consensys [3], Trigno EMG Reports Android App [7], and Phyphox [5] on iOS platforms.

1.1 Motivation

Given the group's awareness of advancements in wearable sensors and ubiquitous computing, an opportunity was seen to discover the complexities of human movement and behaviour. Like many other engineering enthusiasts, the group is driven by the need to bridge the gap between sensor data collection and its practical applications in domains such as healthcare, Human-Computer Interaction (HCI), and, possibly, sports analytics. Understanding the intricacies of full-body motions has great potential for developing human-centric technology, improving wellness monitoring, and optimizing user interfaces for more intuitive engagement.

1.2 Scope and Objectives

The core objective of this report is to provide a comprehensive overview of the team's preparations, obstacles faced, and solutions reached along the route. The report also includes important findings and conclusions from a thorough measurement campaign. A total of 20 tests were meticulously conducted, each involving individual participants, with a primary focus on examining the functionality of sensors, capturing intricate muscle movements, electrical activity of the heart, and accelerometer data.

1.3 Overview of the Report

The report is structured as follows:

- The introduction part provides a contextual foundation for understanding the logic behind the gesture database attempt and its importance in the field of Ambient Intelligence.
- The methodology section then outlines the methodical technique used in sensor selection, setting up, and data collection methods.
- The findings section highlights insights gained from the dataset, including participant demographics, database structure, and preliminary observations.
- An extensive discussion follows, in which the implications of the dataset, its representation, and prospective uses, and then the paper concludes with the proposed future work.

2 Methodology

This methodology section outlines the systematic approach adopted to conduct the study, from sensor selection and placement to the execution of data-gathering procedures, ensuring reliable and meaningful results for analysis; and the reproducibility of the results obtained.

2.1 Employed sensors

The group was provided with the following types of sensor: Shimmer3 ECG Unit with 5 ECGs snap lead wires to record the electrical signals from the heart [3], Trigno Quattro Sensor with 4 mini ECGs and one Inertial Measurement Unit (IMU) to measure electrical currents generated in muscles during its contraction representing neuromuscular activities [4], two smartphones coming with built-in accelerometers, gyroscopes, heart rate monitors, and Global Positioning System (GPS) capabilities, and mm-wave radar. The team found early on equipment complications with the mm-wave radar; and due to lacking support, opted to use EMGs, ECGs, and two smartphones equipped with the Phyphox application to capture accelerometer sensor data.

- ***EMG Sensors:*** EMG sensors from TrignoQuattro [4] were strategically placed on major muscle groups of the participants' dominant arms. Specifically, sensors were positioned on the triceps, deltoid, and two muscles of the forearm, which can be seen in Figure 1, to capture muscle activity during various predefined movements.

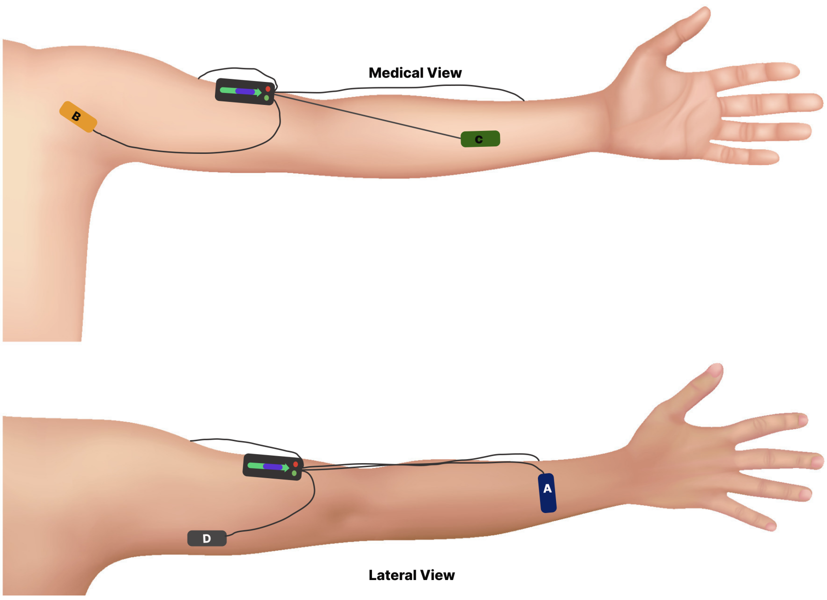


Figure 1: EMG sensors' placements on dominant arm

- ***ECG Sensors:*** ECG sensors from Shimmer were placed on the participants' bodies to record the electrical activity of the heart. Sensors were positioned near the collarbones, hips, and one away from the heart to ensure accurate capture of heart rate and rhythm. Figure 2 gives a better visualization of the sensors' placements. The abbreviations LA, RA, LL, RL, and FH characterize the following anatomical positions: Left Arm, Right Arm, Left Leg, Right Leg, and Far Heart, respectively.
- ***Accelerometers:*** Accelerometer data was collected using both smartphones running the Phyphox application. Smartphones were placed in the participants' trousers pockets to record changes in velocity and body movement with accuracy.

2.2 Data Gathering

The data-gathering process involved meticulous planning and execution of various actions to capture meaningful data from the sensors. The preparation phase included creating advertisements to call for volunteers willing to contribute to scientific research, designing the information that was to be obtained from the participants while

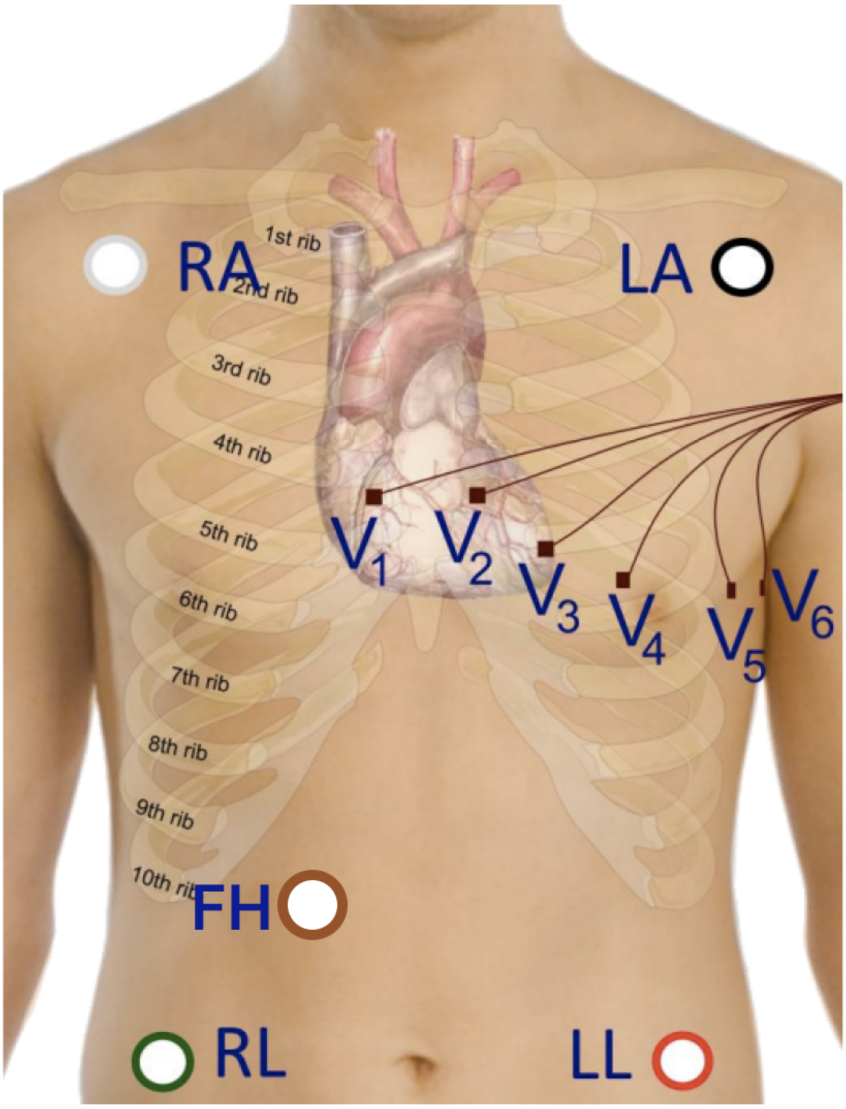


Figure 2: ECG sensors' placements on body

ensuring anonymity, and the actions to be performed by them. Additionally, a consent form was prepared to ensure ethical compliance (available in Appendix B).

2.2.1 Personal data

Volunteers were asked, before participating in the test, to answer a poll (available in Appendix A) requiring information about participants' age, weight, height, dominant arm, and ethnicity. The reasoning behind the need to gather this concrete data has been argued as follows:

- **Dominant arm**: Due to the limited availability of EMG sensors, these will only be equipped on the dominant arm; and the captured behaviours when writing and using a smartphone will necessarily be defined by which arm is the dominant one for the participant.
- **Age**: Different life stages may exhibit unique movement patterns, and this information allows any future usage of this data set to explore and appreciate these variations; while remaining aware of potential age groups that could be left behind in this data collection.
- **Height and Weight**: The participant's height and weight are essential anthropometric measures that can influence body mechanics and movement patterns. This information is valuable for analysing how body proportions may impact potential outcomes of studies done on the gathered data; while being able to recognize potential body types whose data the team might have failed to gather or consider.

- **Gender**: By collecting gender data, the team hopes to be able to explore potential variations in body movement across different gender expressions, fostering a more comprehensive understanding. Any uses of the dataset will also be able to acknowledge potential missing representation in the data.
- **Ethnicity**: Investigating body movement across diverse ethnic backgrounds helps ensure inclusivity and recognizes potential influences on motor behaviour. It allows for a more nuanced analysis of how various populations may differ in their movements; and can help in recognising missing representation in the dataset. While this dataset may not capture the full spectrum of ethnic diversity, this information is essential for future considerations and discussions.

2.2.2 Tested actions

Each test session consisted of specific actions performed by participants, designed to elicit distinct responses from the sensors. These actions were carefully chosen to cover a range of movements and scenarios, allowing for comprehensive data collection.

During this data-gathering phase, participants were instructed to perform a series of predefined actions while wearing the sensor equipment. These actions included writing a paragraph by hand, typing a predefined paragraph on a computer, and scrolling through social media for one minute while sitting. While standing, participants were asked to walk around the room for one minute and perform various activities such as jogging, 3 times squats, clapping, and 12 activities from the MSRC-12 dataset. The MSRC-12 dataset is a compendium of 12 body movements, ranging from lifting arms to punches and kicking, developed by Microsoft and widely used in the literature for neural network training, sensor analysis, etc [1, 2].

The list of activities and their duration and a short description of each are as below:

- **Writing by hand:** Participants were requested to manually write the sentence: "The university is named in honor of Alvar Aalto, a prominent Finnish architect, designer, and alumnus of the former Helsinki University of Technology." There was no time constraint for this task.
- **Writing on a PC:** Participants were directed to type the sentence: "The close collaboration between the scientific, business, and arts communities is intended to foster multidisciplinary education and research." on a computer. This task was also not time-bound.
- **1-min scroll through social media:** Participants were instructed to spend one-minute browsing social media on their mobile phones, which included activities like scrolling through feeds and viewing posts.
- **10-sec stand up:** Participants were asked to maintain an upright standing position for ten seconds.
- **30-sec walk around:** Participants were directed to walk around for thirty seconds.
- **30-sec jogging:** Participants were invited to jog on the spot for thirty seconds.
- **30-sec stand-up rest:** Participants were asked to stand upright and rest for thirty seconds.
- **15-sec 3x squat:** Participants were instructed to perform three squats within fifteen seconds.
- **10-sec clap hands:** Participants were asked to clap their hands for ten seconds.
- **15-sec Lift Arms MSRC-12 movement 1:** Participants were directed to continuously raise their arms overhead for fifteen seconds.
- **15-sec Duck MSRC-12 movement 2:** Participants were asked to mimic a ducking motion for fifteen seconds.
- **15-sec Push Right MSRC-12 movement 3:** Participants were instructed to simulate pushing a chair towards the right side for fifteen seconds.
- **15-sec Look around using Goggles MSRC-12 movement 4:** Participants were asked to pretend to look around while wearing goggles for fifteen seconds.
- **15-sec I've had enough MSRC-12 movement 5:** Participants were asked to mimic a movement indicating frustration or exhaustion for fifteen seconds.
- **15-sec Change Weapon MSRC-12 movement 6:** Participants were asked to simulate the action of changing a weapon for fifteen seconds.

- **15-sec Wind it Up MSRC-12 movement 7:** Participants were instructed to simulate physical movements associated with feeling excited or energized for fifteen seconds.
- **15-sec Shoot MSRC-12 movement 8:**Participants were asked to simulate the action of shooting a firearm for fifteen seconds.
- **15-sec Bow (with joint hands) MSRC-12 movement 9:** Participants were asked to perform a bowing motion with their hands joined together for fifteen seconds.
- **15-sec Throw (over the shoulder) MSRC-12 movement 10:** Participants were asked to simulate the action of throwing an object over their shoulder for fifteen seconds.
- **15-sec Punch (with both hands) MSRC-12 movement 11:** Participants were asked to mimic a punching motion with both hands for fifteen seconds.
- **15-sec Kick MSRC-12 movement 12:** Participants were asked to mimic a kicking motion for fifteen seconds.

Throughout the data-gathering process, attention was paid to environmental factors and potential sources of interference that could affect data accuracy. Participants were instructed to perform actions in controlled settings to minimize external influences on sensor readings.

2.2.3 Data Collection Software

Data from the three sensor types (ECG, EMG, and accelerometers) can be collected in a myriad of ways. To ensure the reproducibility and traceability of potential problems, the software used for the collection is described here, as well as the general processes employed for data export before curation.

- **ECG**: Consensys
  In this study, the team used Consensys V1.6.0 to communicate with Shimmer devices and manage data collected from the sensors. ConsensysPRO is a host side application used to configure one or more Shimmers and stream data from it. Consensys allows users to display and record data received from Shimmer devices streaming over Bluetooth. Users can select the sampling rate, enable/disable specific sensors, enable/disable power monitoring, and change parameters such as the kinematic sensors' sensitivity. Once captured, the data can then be saved in .csv format for further interpretation and analysis [3].

- **EMG**: Trigno EMG Reports Android App
  EMG Reports is an app developed by Delsys, the same company behind the Tringo Quattro sensors. Available for Android devices, the app allows for seamless connection with the sensors via Bluetooth, the display of real-time data, and the recording of such data for later use [7]. The last one was employed for this data collection, which results in a data point being recorded every 15 ms. This data was then was exported to both available formats, .shpf and .xlsx. For data curation purposes, the .xlsx format was used due to it allowing for easier data handling; yet both file types are included in the raw data part of the generated database.

- **Accelerometers**: Phyphox IOs app
  Phyphox is a smartphone application that lets users utilize their smartphone's sensors for physics experiments. It offers a user-friendly interface for accessing and analysing data from various sensors, including magnetometers, gyroscopes, accelerometers, etc (see Figure 3) [5].

  The application has an option "accelerometer sensor without g" in Phyphox. This means that the accelerometer data provided by this option only represents the acceleration experienced by the device in a particular direction, excluding the gravitational acceleration acting on the device [8]. To capture data on the dynamic acceleration of an item or device without the impact of gravity for use in various physics experiments and analyses, the group decided to use the accelerometer sensor without the g option in Phyphox for this project. The software also lets the user easily export the data in many common formats, which are .xlsx and .csv, after every measurement. The data could then be saved and/or shared using any app [9].

Figure 3: Phyphox interface and sensor options [5].

2.3 Data Curation

Following the completion of data gathering sessions, the collected data underwent a thorough curation process to ensure its quality and reliability for further analysis. This involved several steps aimed at cleaning, organizing, and preparing the data for analysis; and most importantly, standardising the format of the data to allow for future joint analyses of the different sensors. With that goal in mind, timestamps were standardised into seconds (s), and sensor measurements for the EMG and ECG sensors to millivolts (mV).

The first step in the data curation process involved inspecting the collected data for any inconsistencies or anomalies. This included checking for missing or corrupted data points and identifying outliers that could potentially skew the analysis results. Any such issues were addressed using data-cleaning techniques, such as interpolation or removal of erroneous data points. Most missing or corrupted data points were detected immediately after each measurement session by comparing the duration of all data graphs or timelines to ensure synchronization. The corruption mainly occurred with the accelerometer data collected by the Phyphox application on iOS phones placed in participants' pockets, where factors such as pocket tightness or size affected the data duration.

Once the data was cleaned, it was organized into a structured format suitable for analysis. This involved converting all time-related data into seconds, as previously mentioned, and labelling the data according to the actions performed by the participants. The curated data was then organized into datasets for each sensor modality. For this step, three timestamps of each participant run were key in successfully splitting the data into segments for each action.

- **Timestamp 1** refers to the moment the participant starts to write by hand.
- **Timestamp 2** refers to the moment the participant starts to write in a computer.
- **Timestamp 3** refers to the moment the participant starts scrolling through social media using their personal smartphone.

Critically, the actions were planned in a way in which everything the participant did once they started scrolling through their smartphone was tightly timed, easing the data labelling done afterwards. This ultimately leaves only two actions of variable length, the writing ones, that need to be carefully specified for each participant.

Finally, the curated data was stored in a GitHub repository for further uses analysis [6]. This repository was designed to facilitate easy retrieval and sharing of the data while ensuring compliance with data privacy and security regulations. The repository, thus, includes the raw data, the curated data, and the code used to label each sensor's data into action-related bits; as well as the relevant gender, ethnic, and anthropometric data of each participant.

2.4 Ethical Considerations

The ethical aspects of the study were carefully addressed to ensure the protection of participants' rights and privacy. Approval was obtained from the Aalto Board; and, as previously stated, all participants were required to provide informed consent before participating in the study (Appendix B). Consent forms outlined the purpose of the research, the procedures involved, and the rights of the participants. Additionally, measures were taken to anonymize the data to protect the privacy of the participants, ensuring that their identities remained confidential throughout the study. In the published database, only the participant codes are included; with no personal data nor contact method other than the data deemed necessary for research purposes (age, gender, ethnicity, and anthropometric data collected in the poll). Added to that, and following Aalto ethical guidelines, the participants have the irrevocable right to remove their participation at any time; meaning their data will be, if requested, completely removed from the resulting database.

Regarding the background data requested in the poll that the participants are asked to fill (available in Appendix A); it does cover some topics that may be sensitive to some, like gender identity or ethnicity. While the need for this data has been deemed justified, as there was a need to be able to recognise potential blind spots in the data due to a potential lack of diversity in the participants, careful consideration was put into how these questions were presented so as to avoid creating an uncomfortable experience for the participants.

For gender, it was gender identity, rather than sex, as it allows for a more comfortable experience for the participant, broader options for self-identification, and more nuanced and grounded data for potential future uses of the database. Besides non-binary, woman, and man, the options "Prefer not to answer" and an open answer "Other" (where the participant can write down their preferred answer) were included; following industry recommendations and guidelines [10].

For ethnicity, relevant literature and standard practices followed by institutions that usually deal with multi-ethnic populations were followed. General rules, typically understood as best practices were thoroughly followed, like using "ethnicity" rather than "race", using terms as broad as possible, geographical references rather than national ones, allowing for multiple choices to be selected, and including the previously mentioned "Prefer not to answer" and open answer "Other". By mixing the practices used in the literature and by multiple polling and census institutions, the following terms were offered to the participants [11, 12, 13, 14, 15]: Black or African American, Asian, Native American or Indigenous, Pacific Islander or Aboriginal, Latina or Latino, Indian, Middle Eastern or Northern African, and White (plus "prefer not to answer" and "others", as previously stated).

Similarly, for the anthropometric data, a self-report was preferred in a search for balance between participant comfort and data needs. Thus, weight and height were asked directly in the poll as questions that only accepted numerical values; while the dominant arm where the EMG sensors were to be put was requested in a select-from-list type of question. The spectrum of arm preference was included in the list to account for potential ambidextrous participants, who were then prompted to choose an arm for the realization of the test.

2.5 Known Issues and Limitations

During the study, despite thorough planning and execution, several issues and limitations arose; which are important to recognise and highlight here for future uses of the resulting data set. None of the issues discussed here seems; upon initial inspection and verification, to have compromised the resulting data set. Nonetheless, depending on the usages this data receives in the future, it is worth checking if these issues and limitations may have an impact on any further conclusions.

First, and most direct result of the employed methodology, is the fact that the participants self-reported key anthropometric data before beginning the test. In a balancing act between data accuracy and participant comfort, neither age, height, nor weight were verified in any way after self-reporting. While the team has no reason to suspect foul play with the data provided by any participant; the fact remains that, as self-reported data, small deviations from the truth are not to be ruled out.

Moving towards the actual data gathering on the sensors, it is worth remarking that varying arm lengths among participants presented challenges in EMG sensor placement due to cable length constraints. This results in inconsistent sensor placement in the participants' arm; which can potentially compromise the data usability. Additionally, participants' inconsistent use of their dominant arm during test sessions resulted in fluctuations in sensor readings and data interpretation. This inconsistent use was observed to varying degrees depending on the participant, but ultimately can be traced to the fact that only one arm was being recorded by the EMG sensors. This created a sense of unbalance in many participants, who hesitated to act normally with that arm; a behaviour that was very visible when comparing the freedom that the other arm tended to enjoy.

Moreover, occasional Consensys software errors occurred during data collection. These have, after data integrity verification, not affected the collected ECG data; but did impair at many points the normal functioning of the test performed by the participants. These differences in the reality experienced by the participants; but the length of this effect in the collected data is hard to quantify. This is in stark contrast with the ease of usage of Trigno Android app or the Phyphox iOS app used for gathering accelerometer data.

On the flip side, it was a hardware limitation what compromised some accelerometer data. Due to the size of participants' clothing pockets, the actual position of the accelerometer in the participant's leg was inconsistent. While this was intended, as the data set aims to reflect the wide range of potential positions a phone can take on someone's pocket; this did result in a few participants data being altogether unusable for some parts of the test due to the accelerometer stopping the recording due to unlucky skin contact.

Lastly, a known limitation exists regarding the inconsistencies in the test environment. Due to university space availability constraints, many tests were performed in different rooms. This resulted in some variations in the test, such as chairs with and without wheels. In turn, this introduced variability that could impact sensor measurements' consistency.

3 Results

The results of the study provide insights into the functionality of the sensors and the challenges encountered during data collection. Analysis of the collected data revealed patterns in sensor readings corresponding to different actions performed by the participants. These results contribute to a better understanding of sensor behaviour and inform future research in the field of ambient intelligence.

3.1 Participants

A total of 21 volunteers answered the team's call for participants. From these, 20 ended up participating; whose data has been labelled P1 - P20 accordingly. The population these people represented can be seen in the following subsections.

3.1.1 Anthropometric Diversity

Due to the fact that all the participants were university students or familiar with the university, the age bracket is quite limited. On the flip side, they did represent remarkable variability in height; while, admittedly, not so much in weight. Still, taking into consideration the population size, the team deems it diverse enough. Relevant statistical data for these values can be consulted in Table 1, and the historiographs to see the actual data distributions are included in Figure 4.

Regarding the dominant arm, almost no variety was achieved. Every participant chose to perform the proposed actions with the right hand, with only one reporting as ambidextrous (the rest self-defined as having the right as the dominant arm).

	Age (years)	**Height** (cm)	**Weight** (Kg)
Maximum	35	184	88
Minimum	22	154	49
Median	27	175,5	67,5
Average	27,55	173,35	68,5
Standard deviation	3,886	8,425	12,626

Table 1: Statistical data of the anthropometrics of the participants (self-reported).

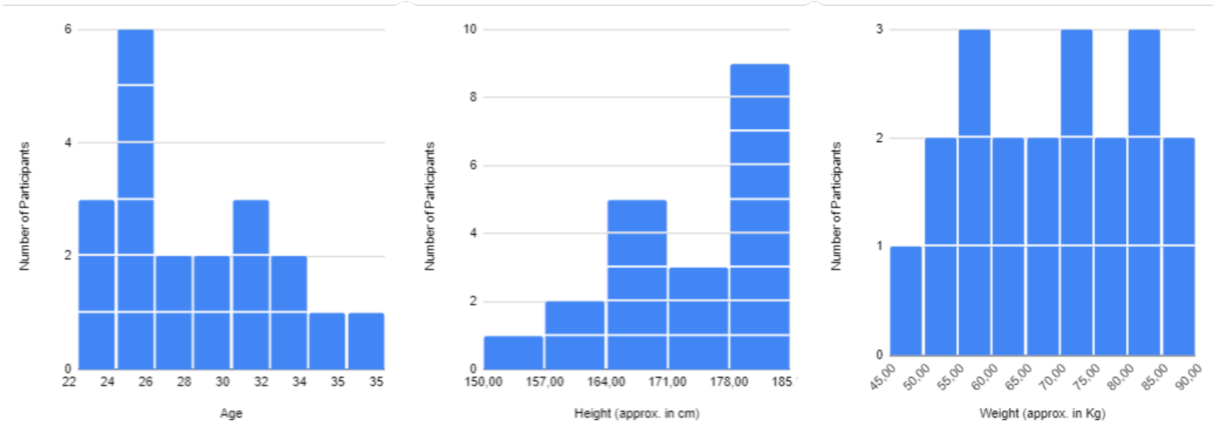


Figure 4: Historiographs for participants' Age, Height, and Weight.

3.1.2 Gender Diversity

Ensuring gender diversity has been a priority for the team; mainly due to the fact that Science, Technology, Engineering, and Mathematics (STEM) university faculties tend to over-represent men. By reaching out to more women, however, a better balance has been achieved when compared to the actual men-to-women ratio of the Computer Science faculty at Aalto University. In total, 14 men, 5 women participated, with one participant preferring not to disclose their gender (see Figure 5).

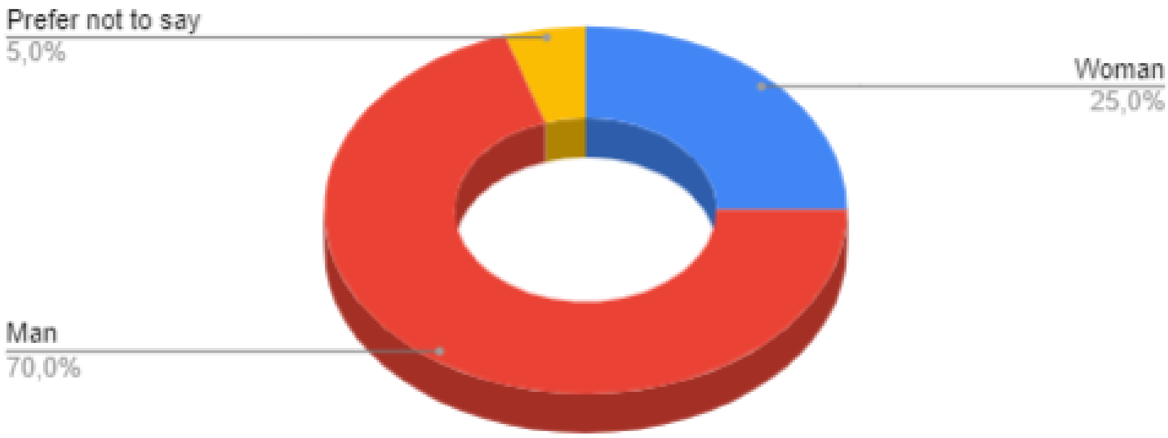


Figure 5: Gender distribution of the participants.

3.1.3 Ethnic Diversity

Following the methodology previously discussed, participants' ethnicity was an open question that allowed for the selection of multiple ethnicities, as well as the option to self-describe or not answer. When aggregating the data, this results in more ethnic backgrounds than there are participants, with Middle Eastern or Northern African coming first with 8 participants, followed by 7 ethnically Asian, 5 White, 1 Latino, 1 Indian, and 1 who did not specify their ethnic background (see Figure 6).

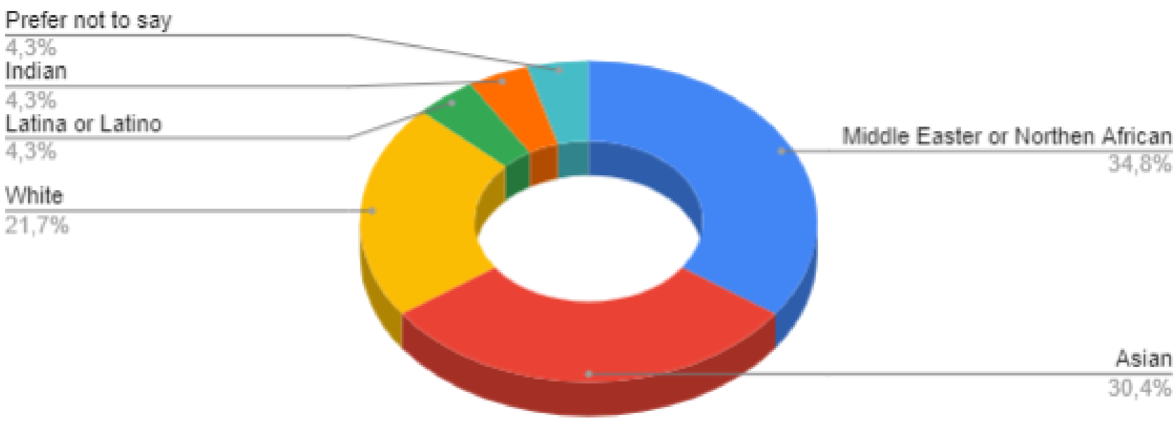


Figure 6: Ethnic distribution of the participants.

3.2 Database

In the course of the measurement campaign, a significant amount of raw data was collected. This data, along with the corresponding data labelling code written in Python, is stored in a dedicated GitHub repository [6]. This approach ensures version control and facilitates collaboration among team members.

The repository is organized into three folders, each corresponding to the data collected from specific sensors used in the measurement. These folders are named as follows:

- shimmer: Contains data from Shimmer sensors and the .py file for data labelling [16].
- phyphox-ios: Stores data from the Phyphox app and the .py file for data labelling [17].
- EMG RMS: Contains data from Trigno sensors and the .py file for data labelling [18].

Each folder contains data files in .csv and/or .xlsx format. Each file corresponds to one participant, with a total of 20 files for 20 participants. This structure allows for easy access and manipulation of individual participant data. While some challenges were encountered when dealing with certain file formats, but these were resolved, ensuring seamless access to all data.

Tables 2, 3, and 4 provide a snapshot of the first five rows of a random file from each directory. This gives a glimpse into the structure and format of the data that is being worked with. Added to that, Figure 7 provides an example illustration of the data available after curation. Figure 7a showcases clear spikes every time P7 performed a "punch"-type action, while Figure 7b shows the differences between scrolling (high spikes) and texting (relative calmness).

X[s]	a: EMG RMS 1-1	X[s]	a: EMG RMS 1-2	X[s]	a: EMG RMS 1-3	X[s]	a: EMG RMS 1-4
445.0049999997162	1.411132e-05	445.0049999997162	2.544453e-05	445.0049999997162	8.968555e-06	445.0049999997162	4.861837e-06
445.01999999971616	1.340064e-05	445.01999999971616	2.522434e-05	445.01999999971616	8.977973e-06	445.01999999971616	4.756387e-06
445.03499999971615	1.334903e-05	445.03499999971615	2.531131e-05	445.03499999971615	8.933933e-06	445.03499999971615	4.806471e-06
445.04999999971614	1.300371e-05	445.04999999971614	2.505006e-05	445.04999999971614	8.629149e-06	445.04999999971614	4.645514e-06
445.0649999997161	1.305453e-05	445.0649999997161	2.558862e-05	445.0649999997161	8.405879e-06	445.0649999997161	4.621192e-06

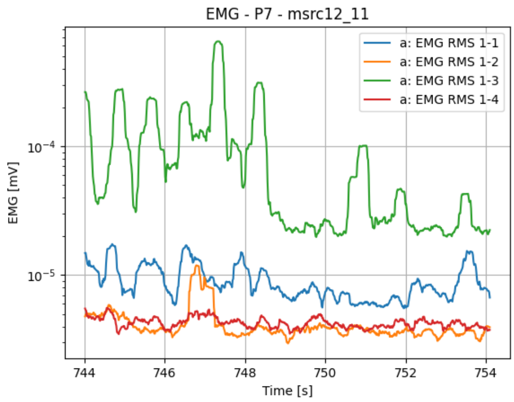
Table 2: Five-row sample of EMG data [18].

Time (s)	Linear Acceleration x (m/s2̂)	Linear Acceleration y (m/s2̂)	Linear Acceleration z (m/s2̂)	Absolute acceleration (m/s2̂)
0.01658866694	-0.8527360037	-1.25051536	0.3484145308	1.553170964
0.02657666709	-0.08654580832	-1.161184231	1.064890237	1.577919583
0.036563667	0.3832710989	-0.8029788297	1.269515303	1.550271215
0.04655166669	0.506487574	-0.5095892295	1.40607527	1.579005545
0.05653866706	0.4875696386	-0.5557474422	1.163010024	1.378104382

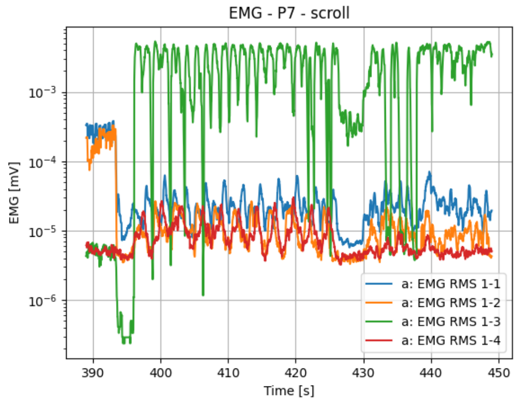
Table 3: Five-row sample of phyphox accelerometer data [17].

Shimmer_C9EB_Timestamp_Unix_CAL	Shimmer_C9EB_Accel_LN_X_CAL	Shimmer_C9EB_Accel_LN_Y_CAL	Shimmer_C9EB_Accel_LN_Z_CAL	Shimmer_C9EB_Accel_WR_X_CAL	Shimmer_C9EB_Accel_WR_Y_CAL
ms	m/(s2̂)	m/(s2̂)	m/(s2̂)	m/(s2̂)	m/(s2̂)
1.71E+12	0.815217391	11.23913043	3.369565217	-2.140035907	9.474566128
1.71E+12	0.8695965217	11.2173913	3.380434783	-2.19880311	9.450628366
1.71E+12	0.760869565	11.2173913	3.336956522	-2.245362059	9.456690604
1.71E+12	0.7608069565	11.19565217	3.336956522	-2.20706164	9.441053262

Table 4: Five-row sample of ECG data [16].



(a) 15 seconds of MSRC-12 movement 11 (punch).



(b) 60 seconds of scrolling through social media.

Figure 7: EMG data snapshot (from P7).

4 Discussion

This section provides a comprehensive discussion on the insights gained from the data, reflects on the diversity of the participants, and explores the potential of the data in training machine learning algorithms. It then concludes with a look at the future directions of this research.

4.1 Insights from the Data

The data collected provides a comprehensive view of full-body gestures across a variety of activities. The inclusion of ECG, EMG, and accelerometer data significantly enriches the dataset, offering a unique opportunity for in-depth analysis and exploration in the field. The chosen sensors are diverse enough to offer a compelling and overall view of the body status in every moment; while the chosen movements reflect wildly different yet common actions and postures useful for an incredible variety of potential purposes.

The accelerometer data, in particular, is of great interest. Accelerometers are commonly found in smartphones, making them a readily accessible source of data for many people. If some of the potential future works of the data described herein are to be ever explored, this usage of smartphone accelerators can open the door to unimaginable every day quality-of-life improvements. The insights gained from this study could potentially be applied to develop applications that leverage accelerometer data to monitor and analyse body movements in real-time. This could be particularly useful in fields such as health and fitness, rehabilitation, and HCI. Since the chosen movements reflect wildly different yet common actions and postures, making them useful for an incredible variety of potential purposes. From understanding the nuances of body language to developing more intuitive user interfaces, the possibilities are vast.

4.2 Representation

One of the key strengths of this study is the diversity of the participants involved. With no over-encompassing gender or ethnicity in the population tested, it offers a wide view of potential movements. Since one of the objectives was to record natural and everyday behaviours, ensuring the inclusion of a wide range of what one can consider natural and intuitive is definitively a key accomplishment.

However, it's important to acknowledge that there are limitations in the participant pool, which may not fully represent the broader population. For instance, no participant contributed with their left hand being the dominant. Gender-wise, only one of the twenty participants expressed themselves out of the binary norm; and ethnic-wise some representation was still lacking from some backgrounds (like black, islander, and Native American peoples, amongst others). Future studies should aim to include a more diverse range of participants to ensure the generalizability of the findings.

4.3 Potential Applications

The potential of this dataset to train machine learning algorithms is immense. Particularly, it can be used to develop models capable of predicting sensory data from missing or malfunctioning sensors using only the available sensors. This could lead to more robust and resilient systems that can maintain functionality even in the face of sensor failures.

Even beyond general machine learning terms, this data set has an unquantifiable potential for real-life applications. From Healthcare and Sports, where the data could be used to develop systems that monitor patient movements and vital signs in real-time, alerting healthcare providers to any significant changes that may indicate a health issue, or allowing for analysis of movement patterns and improve performance; to Virtual Reality applications where the data can enhance the user experience by making interactions more intuitive and immersive.

4.4 Conclusions

The study has successfully collected a rich and diverse dataset that can serve as a valuable resource for researchers and practitioners in the field of ambient intelligence. The data, which includes ECG, EMG, and accelerometer readings, provides a comprehensive view of full-body gestures across a variety of activities. This breadth and depth of data are rarely seen in similar datasets, making it a unique contribution to the field.

Despite encountering several challenges and limitations, the study successfully collected valuable data that can be used for further analysis and research. The findings contribute to the broader understanding of ambient intelligence and provide a foundation for future studies in the field. Additionally, the study highlights the importance of addressing ethical considerations and known limitations in research projects to ensure the integrity and validity of the findings. Overall, it represents a significant step forward in the understanding of HCI and paves the way for future research in this area.

4.5 Future Work

Future work should focus on expanding the dataset with a larger and more diverse participant pool. Additionally, the potential of this dataset to train machine learning models should be explored further, particularly in the context of predicting sensory data from missing or malfunctioning sensors. This could pave the way for the development of more robust ambient intelligence systems. Yet, the potential of this dataset extends beyond just filling in for missing or malfunctioning sensors. Other branches of STEM can potentially make great use of this data. It could be used to develop more intuitive and responsive systems in a variety of fields, from healthcare and sports training to virtual reality and smart homes. All in all, these applications could revolutionize how we interact with technology, making it more seamless and integrated into everyone's daily lives.

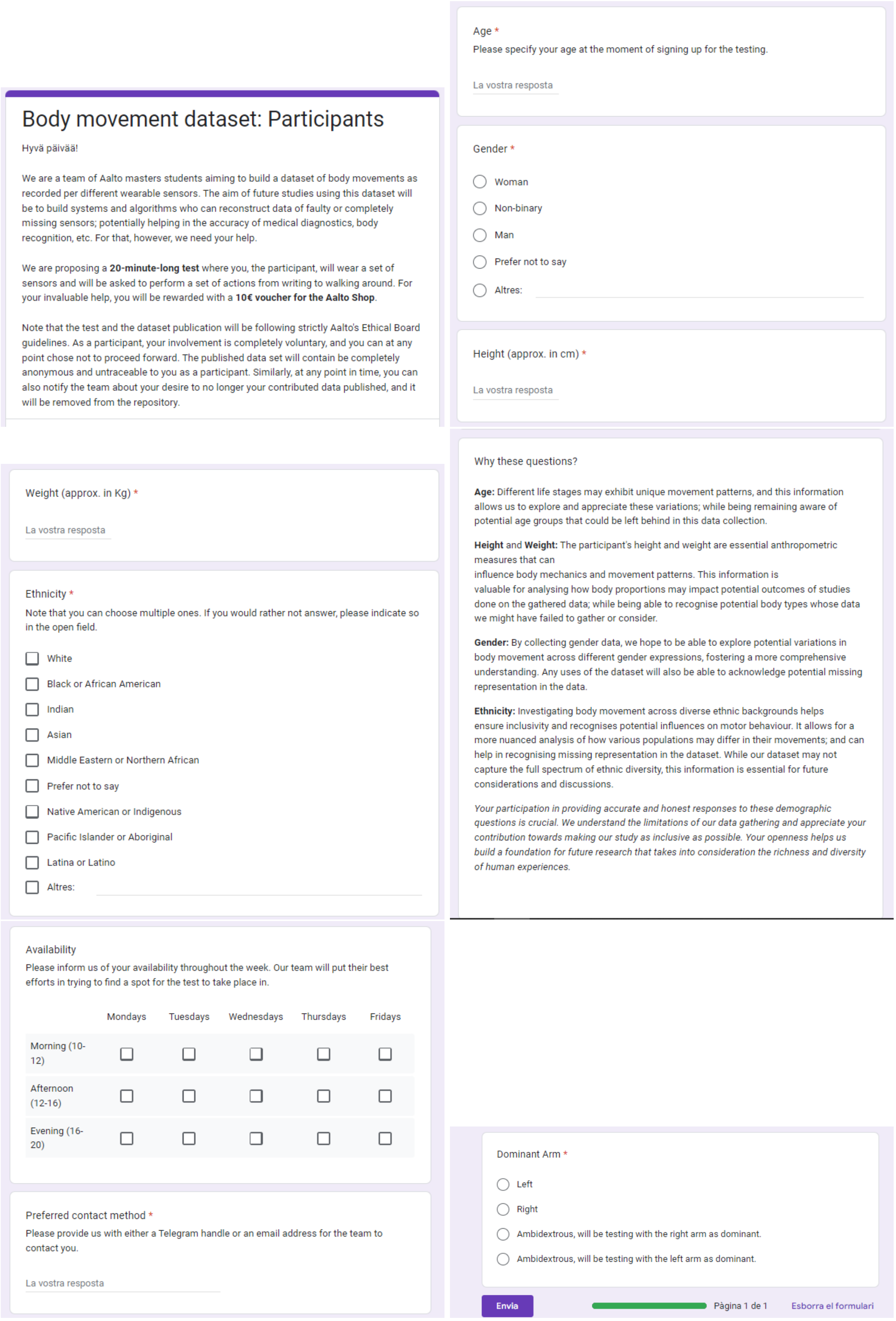
If the data is to be collected again, or the data set expanded, several strategies and considerations can be implemented. Firstly, to mitigate the impact of different arm lengths among participants, standardize the positioning of sensors based on anatomical landmarks rather than relying solely on cable length restrictions. Providing adjustable straps or extension cables can also accommodate variations in arm's length, ensuring consistent sensor placement across participants. Secondly, implement protocols to ensure participants uniformly utilize their dominant arm during test sessions. This can include pre-session instructions and monitoring techniques to encourage consistent movement patterns and minimize variations in sensor readings. Additionally, providing real-time feedback during data collection sessions can help participants maintain consistency in their movements. To solve the problem with the Shimmer program, one could conduct rigorous testing and validation of data collection software to identify and address potential errors before data collection begins.

References

1. PEIXOTO, Júlia Schubert et al. Gesture Recognition using FastDTW and Deep Learning Methods in the MSRC-12 and the NTU RGB+D Databases. *IEEE Latin America Transactions*. 2022, vol. 20, no. 9, pp. 2189–2195. Available from DOI: 10.1109/TLA.2022.9878175.
2. PFITSCHER, Miguel et al. Activity Gesture Recognition on Kinect Sensor Using Convolutional Neural Networks and FastDTW for the MSRC-12 Dataset. In: ABRAHAM, Ajith; CHERUKURI, Aswani Kumar; MELIN, Patricia; GANDHI, Niketa (eds.). *Intelligent Systems Design and Applications*. Cham: Springer International Publishing, 2020, pp. 230–239. ISBN 978-3-030-16657-1.
3. *Shimmer User Manual* [online]. Realtime Technologies Ltd, 2024 [visited on 2024-02-09]. Available from: https://shimmersensing.com/wp-content/docs/support/documentation/Shimmer_User_Manual_rev3p.pdf.
4. *Trigno Wireless Biofeedback System - User's Guide* [online]. Delsys, 2019 [visited on 2024-02-08]. Available from: https://delsys.com/downloads/USERSGUIDE/trigno/quattro-duo-mini-sensors.pdf.
5. *Phyphox - Physical Phone Experiments* [online]. RWTH Aachen University, [n.d.] [visited on 2024-02-19]. Available from: https://phyphox.org/sensors/.
6. MANSOURI, Saeideh; PHAM, Anh; URRUTICOECHEA, Aitor. *Project's Github Repository* [online]. 2024. [visited on 2024-03-22]. Available from: https://github.com/windyinmay/missing-sensors.
7. *EMG Reports Application User's Guide* [online]. Delsys, 2019 [visited on 2024-02-08]. Available from: https://delsyseurope.com/downloads/USERSGUIDE/emg-reports_EN.pdf.
8. *Phyphox Supported Sensors* [online]. RWTH Aachen University, [n.d.] [visited on 2024-02-19]. Available from: https://phyphox.org/sensors/.
9. *Phyphox Supported Export Formats* [online]. RWTH Aachen University, [n.d.] [visited on 2024-02-19]. Available from: https://phyphox.org/export-formats/.
10. MATSUMOTO, Abigail. *How to ask survey questions on sexual orientation and gender identity* [online]. Surveymonkey, 2024 [visited on 2024-03-20]. Available from: https://www.surveymonkey.com/curiosity/ask-survey-questions-sexual-orientation-gender-identity/.
11. *Participant Inclusivity in research | Ethnicity and Race* [online]. University of British Columbia, [n.d.] [visited on 2024-03-20]. Available from: https://ethics.research.ubc.ca/sites/ore.ubc.ca/files/documents/Participant%20inclusivity.Ethnicity%20and%20race.pdf.
12. *Ethnic or cultural origin of person* [online]. Statistics Canada, 2021 [visited on 2024-03-20]. Available from: https://www23.statcan.gc.ca/imdb/p3Var.pl?Function=DEC&Id=1219467.
13. *Ethnic group, national identity and religion* [online]. Office for National Statistics, 2021 [visited on 2024-03-20]. Available from: https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion.
14. LAUX, Richard. *Asking people about their ethnicity - Data in Government* [online]. GOV.UK, 2021 [visited on 2024-03-20]. Available from: https://dataingovernment.blog.gov.uk/2021/06/15/asking-people-about-their-ethnicity/.
15. MATSUMOTO, Abigail. *How to ask ethnicity survey questions* [online]. Surveymonkey, 2024 [visited on 2024-03-20]. Available from: https://www.surveymonkey.com/mp/how-to-ask-ethnicity-questions-survey/.
16. MANSOURI, Saeideh. *ECGs Database* [online]. 2024. [visited on 2024-03-20]. Available from: https://github.com/windyinmay/missing-sensors/tree/main/shimmer.
17. PHAM, Anh. *Accelerometer Database* [online]. 2024. [visited on 2024-03-21]. Available from: https://github.com/windyinmay/missing-sensors/tree/main/phyphox-ios.
18. URRUTICOECHEA, Aitor. *EMGs Database* [online]. 2024. [visited on 2024-03-21]. Available from: https://github.com/windyinmay/missing-sensors/tree/main/EMG%20RMS.
19. *Google Forms* [online]. Alphabet, 2024 [visited on 2024-03-15]. Available from: https://www.google.com/forms/about/.

Appendix A: Data Gathering: Poll

Basic data described in the report was collected using a Google Forms [19]. How this forms looked like can be consulted in Figure 8.



Body movement dataset: Participants

Hyvä päivää!

We are a team of Aalto masters students aiming to build a dataset of body movements as recorded per different wearable sensors. The aim of future studies using this dataset will be to build systems and algorithms who can reconstruct data of faulty or completely missing sensors; potentially helping in the accuracy of medical diagnostics, body recognition, etc. For that, however, we need your help.

We are proposing a **20-minute-long test** where you, the participant, will wear a set of sensors and will be asked to perform a set of actions from writing to walking around. For your invaluable help, you will be rewarded with a **10€ voucher for the Aalto Shop**.

Note that the test and the dataset publication will be following strictly Aalto's Ethical Board guidelines. As a participant, your involvement is completely voluntary, and you can at any point chose not to proceed forward. The published data set will contain be completely anonymous and untraceable to you as a participant. Similarly, at any point in time, you can also notify the team about your desire to no longer your contributed data published, and it will be removed from the repository.

Weight (approx. in Kg) *

La vostra resposta

Ethnicity *

Note that you can choose multiple ones. If you would rather not answer, please indicate so in the open field.

- ☐ White
- ☐ Black or African American
- ☐ Indian
- ☐ Asian
- ☐ Middle Eastern or Northern African
- ☐ Prefer not to say
- ☐ Native American or Indigenous
- ☐ Pacific Islander or Aboriginal
- ☐ Latina or Latino
- ☐ Altres:

Availability

Please inform us of your availability throughout the week. Our team will put their best efforts in trying to find a spot for the test to take place in.

	Mondays	Tuesdays	Wednesdays	Thursdays	Fridays
Morning (10-12)	☐	☐	☐	☐	☐
Afternoon (12-16)	☐	☐	☐	☐	☐
Evening (16-20)	☐	☐	☐	☐	☐

Preferred contact method *

Please provide us with either a Telegram handle or an email address for the team to contact you.

La vostra resposta

Age *

Please specify your age at the moment of signing up for the testing.

La vostra resposta

Gender *

- ○ Woman
- ○ Non-binary
- ○ Man
- ○ Prefer not to say
- ○ Altres:

Height (approx. in cm) *

La vostra resposta

Why these questions?

Age: Different life stages may exhibit unique movement patterns, and this information allows us to explore and appreciate these variations; while being remaining aware of potential age groups that could be left behind in this data collection.

Height and **Weight:** The participant's height and weight are essential anthropometric measures that can influence body mechanics and movement patterns. This information is valuable for analysing how body proportions may impact potential outcomes of studies done on the gathered data; while being able to recognise potential body types whose data we might have failed to gather or consider.

Gender: By collecting gender data, we hope to be able to explore potential variations in body movement across different gender expressions, fostering a more comprehensive understanding. Any uses of the dataset will also be able to acknowledge potential missing representation in the data.

Ethnicity: Investigating body movement across diverse ethnic backgrounds helps ensure inclusivity and recognises potential influences on motor behaviour. It allows for a more nuanced analysis of how various populations may differ in their movements; and can help in recognising missing representation in the dataset. While our dataset may not capture the full spectrum of ethnic diversity, this information is essential for future considerations and discussions.

Your participation in providing accurate and honest responses to these demographic questions is crucial. We understand the limitations of our data gathering and appreciate your contribution towards making our study as inclusive as possible. Your openness helps us build a foundation for future research that takes into consideration the richness and diversity of human experiences.

Dominant Arm *

- ○ Left
- ○ Right
- ○ Ambidextrous, will be testing with the right arm as dominant.
- ○ Ambidextrous, will be testing with the left arm as dominant.

Envia | Pàgina 1 de 1 | Esborra el formulari

Figure 8: Google forms asking the participants on relevant information about themselves.

Appendix B: Data Gathering: Consent form

Before proceeding with the measurements, all participants were provided with the consent form displayed in Figure 9 to ensure they understood the nature of their participation and their rights as volunteer participants in the study.

Participation confirmation
Body Movements Measurement
Data Collection - ELEC-7261 Ambient Intelligence

I have understood that participation is voluntary and at any point in the research study, I am at liberty to notify that I no longer wish to participate in the study, but all the information gathered up until that point is can be used as described in the privacy notice of the research study.

I have received sufficient information about the research study, I have had the possibility to have my questions answered, I have understood the information and I wish to participate in the research study.

____
Signature and name of research participant (choosing to participate can be also expressed for example electronically)

Contact details:
Name:
Phone number:
Email:

Figure 9: Participation consent form.