Biosignal Processing and Activity Modeling for Multimodal Human Activity Recognition

Zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
des Fachbereich 3
der Universität Bremen

vorgelegte

Dissertation

von

Hui Liu

Tag der mündlichen Prüfung:	05.11.2021
Erste Gutachterin:	Prof. Dr.-Ing. Tanja Schultz
Zweiter Gutachter:	Prof. Hugo Filipe Silveira Gamboa

Acknowledgement

I would like to thank my splendid and charismatic doctor mother, Professor Tanja Schultz, the most. Her academic accomplishment, work attitude, and enthusiasm made me determine from the industry to return to campus and regain my desire for academic research and self-improvement. When friends ask me, "What kind of person is your professor?" I always tell them that in her words and deeds, she is not only knowledgeable but also charming. I am honored and proud to be a member of her family of disciples.

And my other academic advisor, Professor Hugo Gamboa, is a person who is profound and renaissance. While I admire him for his academic contributions, I am also deeply attracted by his artistic appeal. His world is full of knowledge and art, and he brings this harmony of rationality and sensibility to work and family life. I still remember the scene where he volunteered to play guitar and his family sang for hundreds of devout believers in the church band. Such wonderful scenes have been accumulated week after week.

My two mentors played the most important roles in the direction of my life. They opened many doors for me that I had never imagined before. Their broad research boundaries made me realize that learning is endless. Their views on research topics have made me refreshing and enlightened. Their help to my scientific research is the source of my persistence. My gratitude to them is from the bottom of my heart.

When I think of my colleagues and friends, there are many unforgettable scenes in my mind. Too many people have helped me in work, in life, and in the field of amateur interest. Let me enumerate them in the chronological order that I met and knew them: my colleagues Dr. Felix Putze, Elke Nakonetzki, Timo Schulze, Dr. Lorenz Diener, Jochen Weiner, Dr. Mazen Salous, Dr. Dennis Künster, Lars Steinert, Moritz Meier, Celeste Mason, Daniel Reich, Eric Schädler, Lisa-Marie Vortmann, Ayimnisagul Ablimit, Miguel Angrick, Marvin Borsdorf, and Kevin Scheck: every one of you reached out in time when I needed help. This is why I like our Cognitive Systems Lab (CSL) family the most. Among colleagues, I especially want to express my great

gratitude to Yale Hartmann, my closest academic comrade. Besides our hand-in-hand cooperation in research, this dissertation would not have been completed so smoothly without him dedicating his spare time to reviewing my drafts and putting forward many opinions. In the research project of *Arthrokinemat*, Dr. Bernd Stetter, Dr. Frieder Kraft, Prof. Thorsten Stein, Prof. Stefan Sell, Florian Görnert, and Prof. Justin Sebastian Lange from other universities or companies gave me great support. Outside of work, my close friends, Dr. Min Weng, Prof. Peijun Xu, Yaming Liu, Dr. Gan Zhou, Zhiguo Zhang, Qian Chen, Huan Song, Dr. Mei Feng, Ira Breitkreuz, Yi Zhang, among others, have given me all kinds of help to make my life meaningful, fulfilling and happy.

Behind everything is a wife who silently supports and gives, with a beautiful, lovely little princess who always likes to watch her father work. A deep kiss to my wife Tingting Xue and my daughter Wanna.

To all of the above mentioned names and others who cannot be enumerated due to space limitations: The grace of dripping water is to be reciprocated by a gushing spring.

Zusammenfassung

Diese Dissertation konzentriert sich auf die systematische Untersuchung der Erkennung menschlicher Aktivitäten (Human Activity Recognition, HAR) und zielt darauf ab, ihre Leistung durch die Entwicklung der sequenziellen Modellierung menschlicher Aktivitäten auf der Grundlage von Hidden-Markov-Modelle-basiertem maschinellem Lernen zu verbessern.

Angetrieben von diesen Zwecken haben wir zunächst eine HAR-Forschungspipeline entwickelt. Basiert auf dieser Pipeline haben wir ein robustes tragbares End-to-End-HAR-System aufgebaut und eine Sensordatenaufzeichnungs- und Aktivitätserkennungssoftware *Activity Signal Kit* (ASK) implementiert.

Dann haben wir mit der selbst implementierten ASK-Software mehrere Datensätze multimodaler Biosignale von über 25 Probanden gesammelt, und einen einfachen Mechanismus zum Segmentieren und Annotieren der Daten implementiert. Mit diesen Daten führen wir umfassende Recherchen zum Offline-HAR-System basierend auf den aufgezeichneten Datensätzen und der Implementierung eines End-to-End-Echtzeit-HAR-Systems durch.

Der wichtigste Beitrag ist, dass wir eine neuartige Aktivitätsmodellierungsmethode für HAR vorschlagen und verifizieren, die die menschliche Aktivität in eine Sequenz von gemeinsamen, bedeutungsvollen und aktivitätsunterscheidenden Zuständen unterteilt, die analog zu Phonemen in der Spracherkennung als *Motion Units* (Bewegungseinheiten) bezeichnet werden.

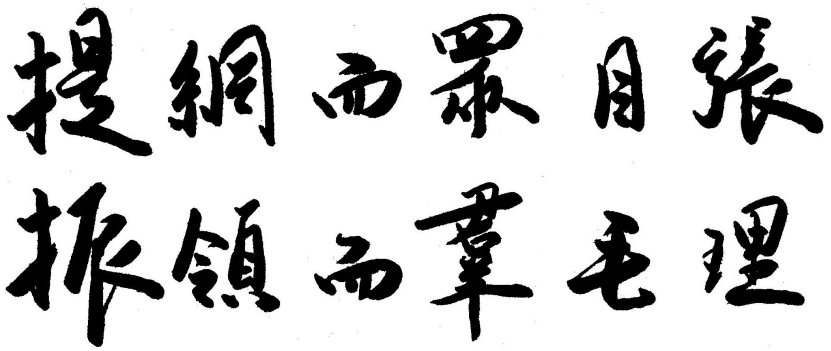
Contents

List of Figures

List of Tables

Chapter 1

Introduction and Motivation



"*Tighten the main rope of the fishing net,*

then all the meshes will be opened widely at a time;

shake the collar of the fur coat,

then all the fur will be straightened out at a blow."[1]

Toqto'a (1314 — 1546), *History of Song.*

[1]The author wrote all calligraphy work in this dissertation and translated them into English (when no appropriate translation exists).

In today's high technology society, *Human Activity Recognition* (HAR) can be used in almost all aspects of life to improve peoples' life quality, such as auxiliary medical care, rehabilitation technology, safety assurance, and interactive entertainment. A suitable algorithm and a high recognition rate is a prerequisite for these applications to work smoothly.

1.1 Definition of Human Activity

Before studying the recognition of human activities, we must first understand what human activity is. Generally speaking, human activity can refer to all behaviors related to human beings, such as brain activity, which do not need to produce any movement. The human activity we studied in this dissertation is in the connotation of kinematics, in which the concept "motion" matches more closely. More specifically, a human activity refers to the movement(s) of one or several parts of the person's body, either atomic or composed of many primitive actions performed in some sequential order (Beddiar et al., 2020). Therefore, HAR aims to label the same activity with the same label even when performed by different persons under different conditions or styles. An HAR system can automatically recognize such human activities using the collected data from various sensors.

As defined above, human activity has a broad denotation. It can refer to a single human motion in a narrow sense, such as "jumping," or a human motion sequence of concurrent, coupled, and sequential motions in a broad sense, such as "cutting a cake," as described in (Gehrig, 2015). Moreover, like most public datasets of human activity, we also categorize some postures which do not produce substantial movement, such as standing and sitting, as the scope of (static) human activity. These activities can also be recognized separately based on motion segments, given suitable equipment.

In this dissertation, the "human activities" we model and recognize are limited to single human motions, including locomotive ones, such as walking and running, and static ones, such as standing and sitting, not involving multiple different primitive motions performed concurrently, coupled, or in some sequential order. Nonetheless, we have already started the modeling and recognition research of concurrent and sequential motions, but it is not the research object of this dissertation.

1.2 Human Activity Recognition (HAR) and Research Purpose

Traditionally, video cameras are widely used in the research and application of HAR systems, making HAR a popular topic in the *Computer Vision* (CV) community. Besides video image signal processing, another core area of the HAR research is time series analysis and human activity modeling based on biosignals. Several well-known time series modeling methods, such as *Hidden Markov Models* (HMMs), *Convolutional Neural Networks*, and *Recurrent Neural Networks*, have shown their abilities in this research field.

In our research, we applied HMMs to build an end-to-end HAR system for assisting the early treatment of gonarthrosis, which is under the framework of the research project *Arthrokinemat* (Arthrokinemat, 2021). Therefore, we used the *GenuTrain* knee bandage (Bauerfeind-GenuTrain, 2021) (see Figure 1.1) provided by one of the project partners *Bauerfeind AG* as a wearable carrier of sensors, aiming to develop an HAR-based mobile technology system that senses its users' movements utilizing proximity sensors.

Figure 1.1 – The Bauerfeind GenuTrain knee bandage applied as a wearable carrier of sensors (Bauerfeind-GenuTrain, 2021).

Figure 1.2 – The integration of the biosignal acquisition devices and sensors into the knee bandage.

Figure 1.3 – Electromyography electrodes' positions on the right leg. Left: musculus vastus medialis and musculus tibialis anterior; right: musculus biceps femoris and musculus gastrocnemius.

Figure 1.2 demonstrates how we integrate biosignal acquisition devices and sensors into the knee bandage, and Figure 1.3 illustrates the position of the electromyography electrodes on the right leg. Using this setup, we focus on raising a practical research pipeline of a wearable HAR system, implementing a baseline offline HAR system, building a real-time recognition software, and researching the activity modeling to improve the recognition performance.

This dissertation is done in the Cognitive Systems Lab at the University of Bremen. Based on the historical contributions to speech recognition under the lab coordinator's expertise, we proposed a hypothesis for more efficiently modeling human activities: Speech and human activities are both time series and have the possibility of being divided into small phases/states. For example, in speech, the phases/states are words, syllables, and phones. What are the correspondent phases/states of human activity? We transfer the existing solutions and technologies in speech recognition to human activity recognition through theoretical and experimental research.

1.3 Contributions

This dissertation aims to research biosignal processing, to improve the modeling of human activities, and to implement a wearable end-to-end HAR system. The major contributions of this dissertation are as follows:

- A novel, HMM-based human activity modeling paradigm — *Motion Units*, a powerful modeling tool to model human activities as intuitive as speech, with good operability, generalizability, and expandability;
- A practical pipeline for HAR research that guides the build of a robust wearable end-to-end HAR system and a series of software developed by following the research pipeline;
- A collection of multimodal biosignal data corpora of human activities from over 25 subjects using the software mentioned above, plus an easy mechanism to segment and annotate the data efficiently;
- A series of studies on the parameters, features, and activity modeling of the offline HAR system based on the recorded data corpora. Our generalized HAR recognizer achieved over 96% person-independent recognition accuracy in a 22-class problem;
- An end-to-end real-time HAR system, which demonstrated its effectiveness and extensibility in international academic occasions.

1.4 Structure of the Dissertation

This dissertation is divided into the following chapters:

Chapter 2 provides background knowledge, including the HAR research pipeline of this dissertation, the biosignal acquisition devices and sensors applied in data collection, the baseline software implemented for dataset acquisition and HAR research, digital signal processing and feature extraction, HMMs, our in-house real-time decoder *BioKIT*, training, recognition, and iterative improvement.

Chapter 3 introduces the sensor integration schema on the knee bandage, three data corpora of human activities from over 25 subjects collected with the software implemented, and an external open-source HAR dataset. All introduced datasets are applied for the research in this dissertation.

Chapter 4 demonstrates the feature dimensionality research experiments and evaluates the performance of the biosignals and the features abstracted from them.

Chapter 5 puts forward different HMM modeling topologies of human activities and the novel, HMM-based human activity modeling paradigm, *Motion Units*, in the analogical study of speech recognition, with good operability, generalizability, and expandability, and presents the experiments and evaluation of the baseline HAR system and the improvement with *Motion Units* to verify the effectivity of *Motion Units* design.

Chapter 6 concludes the dissertation with a summary of results and proposes potential future research directions.

Some sections in the following chapters are revised and improved from the author's previous publications, as listed in Table 1.1.

Table 1.1 – Sections based on the revision and improvement of the author's publications.

Publication	Authorship Position	Sections
(Liu and Schultz, 2018)	1st author	2.3.1, 3.3 and 5.1.2
(Liu and Schultz, 2019)	1st author	5.1.3 and 5.2.2
(Hartmann et al., 2020)	2nd author	4.1.1 and 4.1.2
(Hartmann et al., 2021)	2nd author	4.1.4, 4.2.2, 4.3, and 4.4
(Liu et al., 2021)	1st author	5.2.3, 5.3, 5.5, and 5.7

Chapter 2

Background and Methodology

萬變不離其宗

"*The methods may vary in different applications,*

but the principle is always the same."

Xun Kuang (c. 310 BC — c. 235 BC), *Xunzi*.

HAR research involves various aspects of knowledge, such as hardware (signal acquisition equipment, sensors, among others), software (data collection, archiving, visualization, processing, among others), and machine learning approaches.

2.1 HAR Research Methodology

In this digital age, HAR has been playing an increasingly important role. HAR is often associated to the process of determining and naming human activities using sensory observations (Weinland et al., 2011).

2.1.1 Approaches of Recognizing Human Activities

The recognition of human activities has been approached in two different ways, namely using external and wearable (internal) sensors (Lara and Labrador, 2012). In the former, the devices are fixed in predetermined points of interest, so the inference of activities entirely depends on the voluntary interaction of the users with the sensors. In the latter, the devices are attached to the user, which leads to the research topic of wearable sensor-based activity recognition.

Intelligent homes (Van Kasteren et al., 2010), (Yang et al., 2011), (Tolstikov et al., 2011) are a typical HAR approach of external sensing. Besides, video cameras are also widely used external sensors for HAR system (Mu et al., 2016), (Wang et al., 2016b), (Takeda et al., 2019), (Siddiqi et al., 2014). Among various classification techniques on video-based HAR, two main questions arise: "What action?" (i.e., the recognition problem) and "Where in the video?" (i.e., the localization problem) (Vrigkas et al., 2015). The video-based HAR system tries to answer these two questions based on the recorded video signals. More details about video-based HAR can be found in (Ke et al., 2013).

However, the HAR system based on video camera data has shortcomings, such as privacy and pervasiveness problems (Lara and Labrador, 2012). Although HAR facilitates everyone's life, not every user wants to be "watched" all the time. Moreover, it is not easy to use video technology to obtain the user's entire body image in daily life anytime and anywhere, not to mention the expensive processing and computing technology. These limitations hinder a real-time HAR system from being scalable and widely applicable, motivating the HAR research based on biosignals acquired from wearable sensors.

Wearable sensor-based activity recognition seeks profound high-level knowledge about human activities from multitudes of low-level sensor readings

(Wang et al., 2019). A large body of research involves recognizing various kinds of everyday human activities using different types of sensors, which will be introduced in Section 2.2.

2.1.2 Types of Activities Recognized by State-of-the-Art HAR Systems

From the literature (Lara and Labrador, 2012), seven groups of activities can be distinguished by HAR systems, which are summarized in Table 2.1.

Table 2.1 – Types of activities recognized by state-of-the-art HAR systems (Lara and Labrador, 2012).

Group	Activities
Ambulation	Walking, running, sitting, standing still, lying, climbing, stairs, descending stairs, riding escalator, and riding elevator.
Transportation	Riding a bus, cycling, and driving.
Phone usage	Text messaging and making a call.
Daily activities	Eating, drinking, working at the PC, watching TV, reading, brushing teeth, stretching, scrubbing, and vacuuming.
Exercise/fitness	Rowing, lifting weights, spinning, Nordic walking, and doing push-ups.
Military	Crawling, kneeling, situation assessment, and opening a door.
Upper body	Chewing, speaking, swallowing, sighing, and moving the head.

The design of HAR systems depends on the activities to be recognized, i.e., changing the activity set immediately turns a given HAR problem into a completely different problem (Lara and Labrador, 2012).

In this dissertation, our focus is the generalized human activity modeling and recognition research of the **ambulation** activities in the sense of sensor-based HAR, which are also key aspects of many HAR research pieces and public datasets, such as the *UniMiB SHAR* dataset (Micucci et al., 2017) and the *Enabl3S* dataset (Hu et al., 2018). The reason we study only the ambulation activities in our research will be explained in Section 3.1. However, the research pipeline and the activity modeling method proposed in this dissertation should work not only for ambulation activities. The original purpose is to provide

an activity modeling method for all activity groups, and the research on ambulation activities is the first step because we started under the framework of the *Arthrokinemat* project (see Section 1.2). We are going to promote our method to other activity groups, such as transportation, exercise, and upper body, in the future research work.

2.1.3 HAR Research Pipeline

For researchers in the field of HAR, several key aspects that need to be studied are already established. However, few documents summarize and form a paradigm for the entire framework of HAR research. This may be due to the fact that most researchers focus on the research in one or several fields of HAR, such as modeling optimization, automatic segmentation, feature comparison, scenes of application, among others, rather than the overall HAR process. Given this, we design a practical pipeline, composed of nine topics, to guide our HAR research and clearly indicate the sequence and relationship between each task. Figure 2.1 illustrates the proposed pipeline of our HAR research.

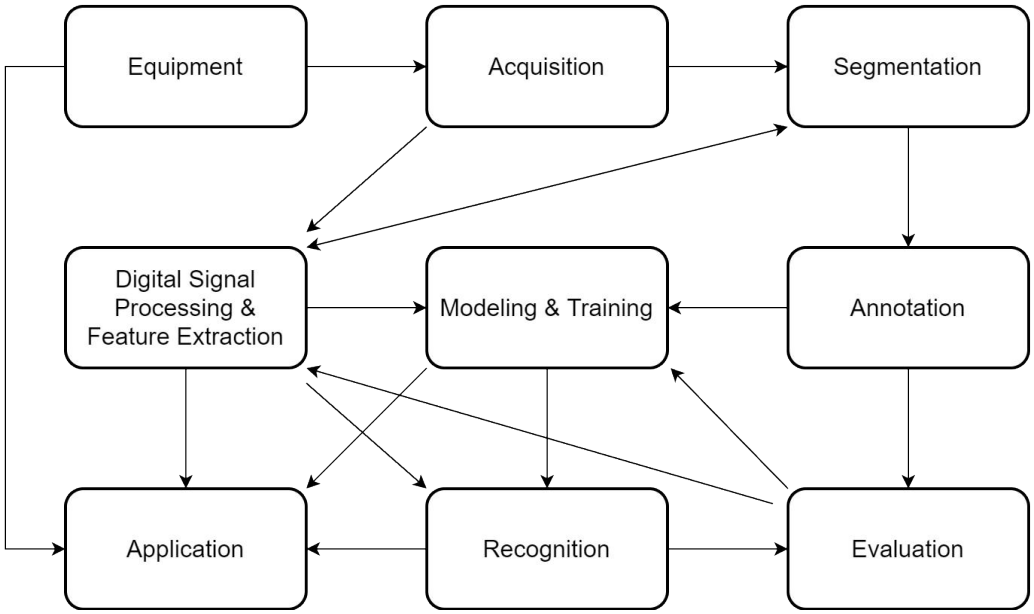


Figure 2.1 – The HAR research pipeline of this dissertation.

The arrows between each task in Figure 2.1 indicate the processing order in the study. These nine research topics are essential and indispensable for our complete end-to-end HAR system. We are working on publishing the HAR

research pipeline with an increment (the segmentation-study-loop), hoping to provide research references to the researchers in the same or similar fields.

2.2 Biodevices and Sensors for HAR

The research object of this dissertation is biosignal-based HAR using wearable sensors. Therefore, selecting the appropriate appliance for biosignal acquisition is essential during our research preparation. There are many considerations for choosing the applicable equipment, such as application occasions, research requirements, and financial conditions. Almost all kinds of biosignal acquisition equipment are capable of particular HAR tasks, depending on different research purposes. For example, when researching daily life assistance or interactive entertainment, placing the mobile phone in a certain pocket of the clothes/trousers to sense inertial signals has become a convenient, efficient, and reasonable design that fits the ultimate use case.

2.2.1 Devices for Multi-Sensorial Data Acquisition

In our research preparation stage, we tested a product of inertial motion unit called *Meta Motion R* (MMR) (see Figure 2.2) by *mbientlab* (MetaMotionR, 2021). It integrates a triaxial accelerometer, a triaxial gyroscope, and a triaxial magnetometer. Moreover, a barometer is also contained.

Figure 2.2 – The *Meta Motion R* with the 3D-printed shell.

We need to place multiple inertial sensors at different body positions in our HAR research under the framework of the *Arthrokinemat* project (see Section 1.2). Although MMR is a comprehensive wearable inertial sensor solution, the synchronization of two MMRs manually or automatically did not soundly succeed after multiple experiments. Besides, it does not provide interfaces to connect more sensors. Therefore, we turned to another recording device,

biosignalsplux (biosignalsplux, 2021), providing expandable solutions of hot-swappable sensors and automatic synchronization, which already existed in our laboratory when we started this dissertation's research,

Some in-house preliminary research work, such as (Rebelo et al., 2013) and (Palyafári, 2015), obtained results that confirm the applicability and robustness of *biosignalsplux* recording devices and its accelerometers, EMG sensors, and goniometer in HAR research. Furthermore, these devices also stably provided high-quality biosignals in our pilot data collection and modeling research (see Sections 3.3 and 5.1.2). Hence, in order to save time and financial cost, we have been using these devices in subsequent research. In the follow-up research, we also selected several additional sensors considered potentially effective for HAR to join the experiments: two gyroscopes, a force sensor, an airborne microphone, and a piezoelectric microphone (in the form of the respiration sensor).

One hub from the *biosignalsplux* research kit (see Figure 2.3) records biosignals from 8 channels, each up to 16 bits, simultaneously. Accelerometer, gyroscope, electrogoniometer, and force sensor signals are relatively slow signals, while the nature of the EMG and both microphone signals require higher sampling rates. Low-sampled channels at 100 Hz are up-sampled to 1000 Hz to be synchronized and aligned with high-sampled channels.

Figure 2.3 – The *biosignalsplux* researcher kit (biosignalsplux, 2021). Left: a complete suite; right: one hub in the kit.

In the three data corpora of different recording periods, we used two or three hubs to acquire biosignal data. We connected the hubs via the synchronization kit (Synchronization-Kit, 2021) that connect the hubs and synchronize all

channels automatically between the hubs at the beginning of each recording session, which ensured the synchronicity during the entire recording. The synchronization kit contains a synchronization cable (Synchronization-Cable, 2021), a multi-sync-splitter (Multi-Sync-Splitter, 2021), a handheld switch (pushbutton) (Handheld-Switch, 2021), and a *Light-Emitting Diode* (LED) (LUX, 2021). The pushbutton was used to realize the novel mechanism of on-recording segmentation and annotation, introduced in Sections 2.4.

Figure 2.4 illustrates all the relevant sensors in this dissertation's research and the synchronization kit. The following sections will introduce the basic knowledge and the related literature of these sensors' application in HAR research.

Figure 2.4 – Sensors and the synchronization kit in the *biosignalsplux* researcher kit. ① triaxial accelerometer (ACC, 2021); ② bipolar surface EMG sensor (EMG, 2021); ③ ground cable for EMG recording (Ground-Cable, 2021); ④ biaxial electrogoniometer (Goniometer, 2021); ⑤ piezoelectric microphone (PZT, 2021); ⑥ force sensor (FSR, 2021); ⑦ triaxial gyroscope; ⑧ airborne microphone; ⑨ synchronization kit (Synchronization-Kit, 2021).

2.2.2 Inertial Sensors

Inertial sensors detect and measure the physical quantities of acceleration, tilt, shock, vibration, rotation, and multiple *Degrees of freedom* (DOF) motion. Inertial sensors mainly include accelerometers (linear acceleration sensors) and gyroscopes (angular acceleration sensors). A *Micro-Electro-Mechanical*

System (MEMS) integrating the monoaxial, biaxial, or triaxial combination of the two sensor types is called an *Inertial Motion Unit* (IMU). Furthermore, An *Attitude Heading Reference System* (AHRS) adds a magnetometer to the IMU so that it can provide additional attitude information including raw, pitch, and yaw.

A large variety of biosignals are captured by multiple inertial sensors to research HAR. As example, (Mathie et al., 2003) applied wearable triaxial accelerometers attached to the waist to distinguish between rest (sit) and active states (sit-to-stand, stand-to-sit, and walk). Five biaxial accelerometers were used in (Bao and Intille, 2004) to recognize daily activities such as walking, riding the escalator, and folding laundry. In (Kwapisz et al., 2010), the authors placed an *Android* cell phone with a simple accelerometer into the subjects' pocket and discriminated activities like walking, climbing, sitting, standing, and jogging. Moreover, (De Leonardi et al., 2018) compared the recognition performance of five classifiers based on machine learning (K-Nearest Neighbors, Feedforward Neural Networks, Support Vector Machines, Naïve Bayes, and Decision Trees) and analyzed the advantages and disadvantages of their implementation onto a wearable and real-time HAR system.

Although inertial sensors are considered in many works, such as the above-listed, performing powerfully for biosignal-based HAR systems, in most of the textbooks of biosignals, like (Dey and Ashour, 2016), (Akay, 2000), (Theis and Meyer-Bäse, 2010), (Wang, 2020), (Naït-Ali, 2009), (Hintermüller, 2016), (Liang et al., 2012), and the signals of the inertial sensors have not been introduced in detail, sometimes even not been mentioned in the context of biosignals, probably because the inertial sensors provide not the bioelectrical signal but the physical, kinetic signal in electrical form. If these signals are utilized to describe the inertial physical quantities of organisms, such as for HAR systems, we classify them as **bio**signals.

Inertial sensors have been widely used in all aspects of our lives, which perceive the kinetic biosignals emitted by our body at all times. The application of inertial information on intelligent solutions facilitates people's lives. In the applicable aspects of navigation (Woodman, 2007), (Veth, 2006), (Veth and Raquet, 2006), (Groves, 2015), (Park and Suh, 2010), human-computer communication (Raya et al., 2012), (Ancans et al., 2017), interactive entertainment (Jung and Cha, 2010), (Wu et al., 2010), (Kim et al., 2019), (Zok, 2014), security system (Fischer et al., 2012), (Eliasson, 2017), healthcare and rehabilitation (Ejupi et al., 2016), (Bo et al., 2011), among others, inertial sensors play pivotal roles. For example, in smartphones nowadays, inertial sensors, including accelerometers, gyroscopes, and even magnetometers, are

usually already integrated on the mainboard, with which recognition systems of human activity (Sousa et al., 2017), (Sousa Lima et al., 2019), (Wang et al., 2016a), (Nurhanim et al., 2017), gesture (Wang et al., 2012), or gait (Sprager and Juric, 2015) can be implemented directly. In many modern game consoles, inertial sensors are also integrated into the gamepads and joysticks to sense common daily activities such as shaking hands up and down, swinging the club, steering the wheel, jumping, shuffling, dancing, among others, to interact with the game program.

2.2.3 Electromyography Sensors

The sensors used to sense the bioelectrical signals emitted by the human body are called electrodes. Electrodes applied to different body positions monitor different biosignals, such as electroencephalogram (EEG), electrocorticogram (ECoG), electrocardiogram (ECG), electrooculogram (EOG), electromyogram (EMG), among others. Although these biosignals are all sensed and transmitted by the sensors with the same name as "electrodes," their characteristics, such as amplitude and frequency, are quite different, determining the requirement to use different amplification, filtering, and preprocessing techniques.

Theoretically speaking, all the biosignals listed above can be applied for activity recognition — when people do different actions, EEG, ECoG, ECG, or even EOG will produce recognizable different signals. If we consider non-kinematic behaviors such as reading, calculation, and other brain activities as "human activities," biosignals such as EEG, EOG, and ECoG can also be greatly helpful in recognizing these activities. However, as specified in Section 1.1, the research object of this dissertation is human activity in the kinematics sense, in which EEG, ECoG, ECG, and EOG do not have pronounced typical characteristics for recognizing simple human activities. In other words, currently, there is no evidence that these four biosignals can be used to distinguish human daily kinematic activities. Only EMG has already been widely used in HAR research because it senses the bioelectrical signals emitted by muscles, and muscle activity is an essential part of most human activities.

The EMG biosignal is the superposition of *Motor Unit Action Potentials* (MUAP) from many muscle fibers in time and space. Surface EMG (SEMG) is a comprehensive effect of superficial muscle EMG and motor unit electrical activity on the skin surface, reflecting neuromuscular activity to a certain extent. Compared to EMG with needle electrodes, SEMG is non-invasive

and much more operable, which is certainly our research choice. All "EMG" mentioned in the following text of this dissertation refers to SEMG.

A grounding signal is usually not necessary for the data collection of bipolar EMG. However, since we used four bipolar EMG sensors simultaneously in the data acquisition, we still applied an extra grounding cable with an electrode (see Figure 2.4③), as stated in the manufacturer's recommendation.

Several works on EMG show that it can be applied to many aspects of the recognition system, such as human movement (Phinyomark et al., 2010), gesture (Zhang et al., 2011), even speech (Jou et al., 2007), (Wand and Schultz, 2014), (Janke and Diener, 2017). EMG also provides the option to predict a person's motion intention prior to actually moving a joint, such as investigated in (Fleischer and Reinicke, 2005) for an actuated orthosis.

2.2.4 Electrogoniometer

A goniometer can provide biosignals by measuring the bending angle of various positions of the human body, such as elbow (the angle between upper and lower arm), knee (the angle between thigh and shank), blade bone (the angle between shoulder and upper arm), and ankle (the angle between shank and foot). According to our pilot research (see Section 5.1.2), the goniometer can help distinguish between some activities to a certain extent, such as standing and sitting. These relatively stable activities are difficult to distinguish by inertial biosignals, but the angle between the thigh and shank will be significant to tell the difference.

Some researchers like (Rowe et al., 2000) and (Sutherland, 2002) applied electrogoniometers to study kinematics. There is not much literature applying goniometers for human activity recognition, except that (Rebelo et al., 2013) and (Palyafári, 2015) studied the classification of human activities based on accelerometers, EMG sensors, and the goniometer attached to the knee. Nevertheless, some scientific research focused on other bio-recognition aspects based on the goniometer, such as gesture (Han et al., 2019) and body-tracking (Carbonaro et al., 2014).

2.2.5 Other Potential Sensors and Sensor Combination

Setups with other sensors for kinematics study have included magnetometers and barometers (Hellmers et al., 2018), or piezoelectric microphones (Teague

et al., 2016). Moreover, (Lukowicz et al., 2004) combined accelerometers with airborne microphones to include a simple auditory scene analysis.

Since we chose the *biosignalsplux* researcher kit as the recording device, we selected several additional sensors considered potentially effective for HAR to join the early experiments: a force sensor, an airborne microphone, and a piezoelectric microphone (see Figure 2.4⑤⑥⑧). Whether these sensors contribute to our HAR research will be discussed in Sections 4.1.3 and 4.2.1.

The majority of previous sensor-signal studies, such as listed in Sections 2.2.2— 2.2.4, are limited to one type of sensor. However, the combination of sensors and the fusion of biosignals may improve the system's robustness and recognition accuracy. (Rebelo et al., 2013) studied the classification of isolated human activities based on accelerometers, EMG sensors, and the goniometer attached to the knee. They successfully recognized seven types of activities, i.e., sit, stand, sit down, stand up, walk, ascend, and descend, with an accuracy of about 98% in a person-dependent recognition. While these results are very encouraging, it remains challenging to robustly recognize the large variety of human everyday activities in the real world.

2.3 Software and Data Acquisition

After selecting the appropriate biosignal acquisition equipment, the next issue is how to collect data. Usually, drivers for different programming languages are provided with the biosignal device sold on the industrial market, allowing users to access the device and record the biosignals themselves.

2.3.1 Activity Signals Kit (ASK): Baseline Software

We developed a software called *Activity Signal Kit* (ASK) with a *Graphical User Interface* (GUI) (see Figure 2.12) and multi-functionalities using the driver library provided by *Plux*. The ASK software connects and synchronizes recording devices automatically. In this dissertation's work, we used two or three *biosignalsplux* hubs (see Figure 2.3) as recording devices in different data acquisition tasks (see Sections 3.3— 3.5). Therefore, ASK collects up to 24-channel sensor data from all hubs simultaneously and continuously. All recorded data are archived orderly with dates and timestamps for subsequent application (Liu and Schultz, 2018).

A novel protocol-for-pushbutton mechanism of segmentation and annotation has been implemented in the ASK software, which will be introduced in Section 2.4. Moreover, the baseline ASK software also provides the functionalities of

signal processing, feature extraction, modeling, training, and recognition by applying our in-house developed HMM-based decoder *BioKIT* (see Section 2.6.4). Since these functionalities are more related to the modeling research and experiments, Chapters 4 and 5 will elaborate on them.

As a summary, the ASK baseline software has the following features:

- Connects to wearable biosignal recording devices smoothly;
- Enables multi-sensorial data acquisition and archiving;
- Implements protocol-for-pushbutton mechanism of practical segmentation and annotation;
- Provides signal processing and feature extraction functionalities;
- Facilitates modeling research with the iteration of training-recognition-evaluation by applying *BioKIT*.

A series of upgraded and expanded versions of the ASK baseline software, such as the real-time end-to-end HAR system and its on-the-fly add-on, have been developed based on modeling and recognition achievements, which will be presented in Section 5.1.3 after the proof-of-concept modeling experiments.

Furthermore, together with a bachelor student, we implemented a *mixture reality* (MR) rehabilitation training game suite, *Activity Game Engine* (AGE). The AGE gaming system has been preliminarily demonstrated at the *Bremen.AI* event (BREMEN.AI, 2021) 2019 and the tutorial section of the 12^{th} International Joint Conference on Biomedical Engineering Systems and Technologies (Liu, 2019). The future work is to integrate the real-time HAR system into the MR game.

2.3.2 Activity Signal Kit MobilE (ASKME): Android Mobile Phone Application

Activity Signal Kit MobilE (ASKME), developed together with a student with his Bachelor thesis (Urban, 2019), is a mobile phone application that implements the same functionalities of device connection, biosignal recording, and data archiving in the mobile phone as the PC's ASK baseline software. Moreover, it provides different GUIs for users and experts and reserves program interface for future medical applications development based on the HAR system. For instance, a traffic light signal is used to indicate knee overload during daily activities.

Figures 2.5—2.8 illustrate its interface on the mobile phone.

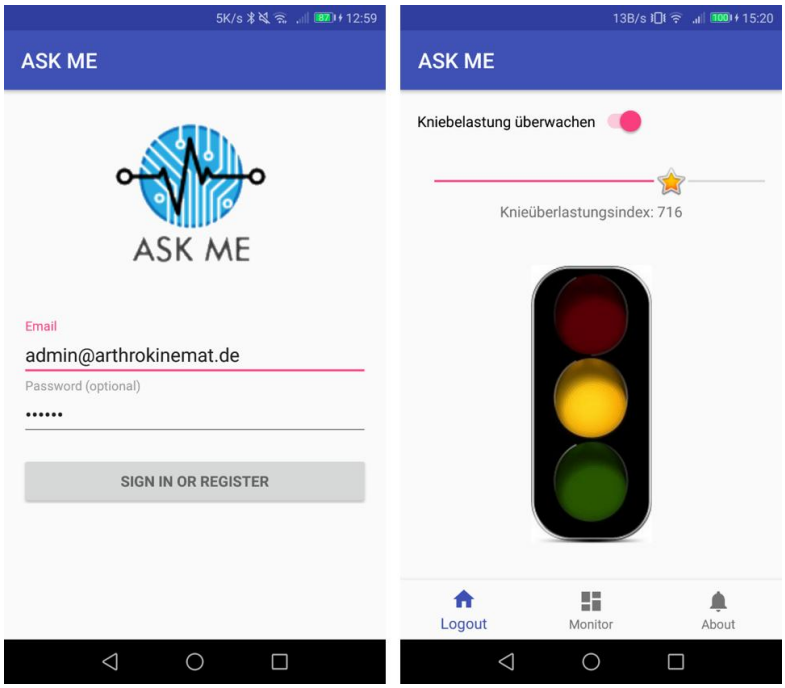


Figure 2.5 – Screenshots of the ASKME mobile phone application: user interface. Left: login; right: patient assistance.

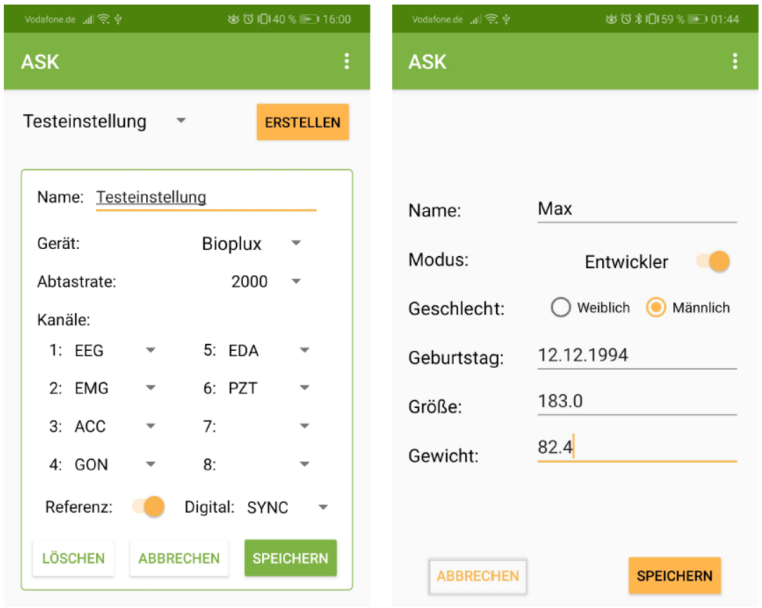


Figure 2.6 – Screenshots of the ASKME mobile phone application: expert interface. Left: sensor setup; right: device setup (Urban, 2019).

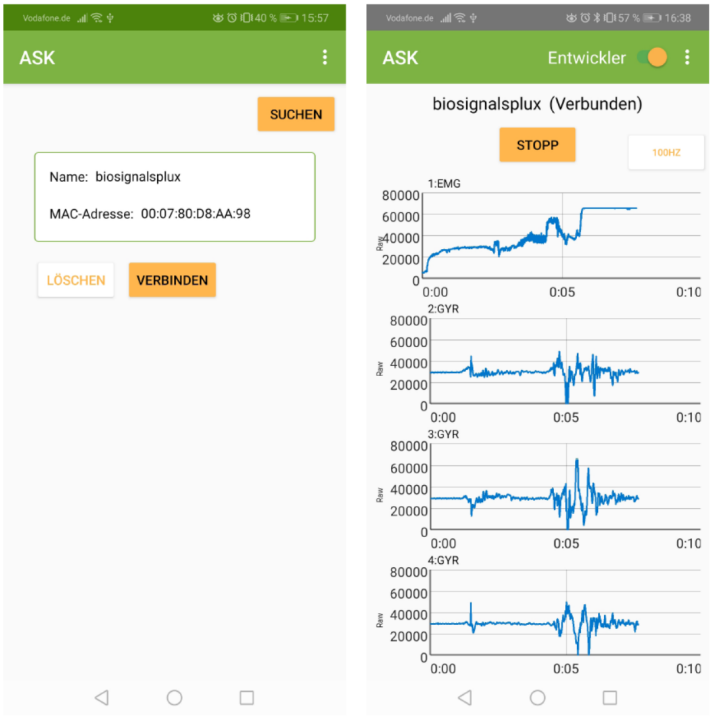


Figure 2.7 – Screenshots of the ASKME mobile phone application: expert interface. Left: device connection; right: data acquisition (Urban, 2019).

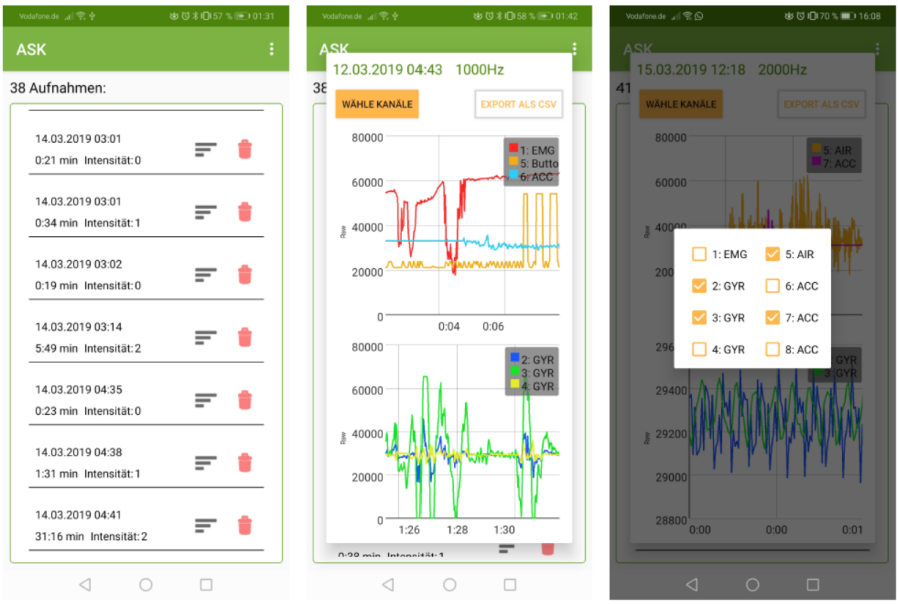


Figure 2.8 – Screenshots of the ASKME mobile phone application: expert interface. Left: file manager; right: signal visualization (Urban, 2019).

2.4 Segmentation and Annotation

The task of segmentation in HAR research is to split a relatively long sequence of activities into several segments of single activity, which are the Tokens used as the smallest identification unit in our in-house HMM-based decoder *BioKIT* (see Section 2.6.4), while annotation is the process of labeling each segment, such as "walk," "run," and "stand-to-sit." In many cases, segmentation and annotation are actually performed simultaneously. As our HAR research pipeline (see Figure 2.1) shows, segmentation is undoubtedly a prerequisite for annotation, and its output will be the input for digital signal processing and feature extraction, while annotation generates labels for two follow-up tasks: training and evaluation.

Segmentation can be performed manually. In video-based segmentation, the biosignal collection is supplemented by video camera(s) recording the whole process. Then, relying on the video, the acquired dataset will be segmented by a dedicated person (Rebelo et al., 2013), (Palyafári, 2015). Another approach of manual segmentation is the use of data visualization. If the collected signals have good recognizable discrimination, we can also segment the data by directly visualizing the signals. Such being the case, video is not necessarily required. If we thoroughly know what happened during the data collection, e.g., through detailed text records, the process will be more efficient. Taking Figure 2.9 as an example, since the sensors are marked clearly in the visualization, if we accurately get informed what activities happened during the data acquisition, we could try to segment and annotate the data manually only based on the data visualization, without applying any video information.

The advantages of manual segmentation are apparent. It is straightforward and intuitive, and the result should be close to the human's understanding of "activity." The segmented data has, therefore, strong rationality and readability. Moreover, manual segmentation can often be accompanied by annotation — marking each segment with a predefined label. However, the shortcomings of manual segmentation cannot be ignored. First of all, manual segmentation has high requirements for the operators: concentration, patience, attentiveness, and even the need to receive some training in advance to adapt to the segmentation requirements for specific research tasks. Even so, manual segmentation is still unavoidably subjective, resulting in poor repeatability and errors due to human factors, which also causes an "intercoder agreement" problem of two or more people (Artstein and Poesio, 2008). (Kahol et al., 2003) used a comparative example (Figure 2.10) to corroborate the subjectivity during the manual segmentation.

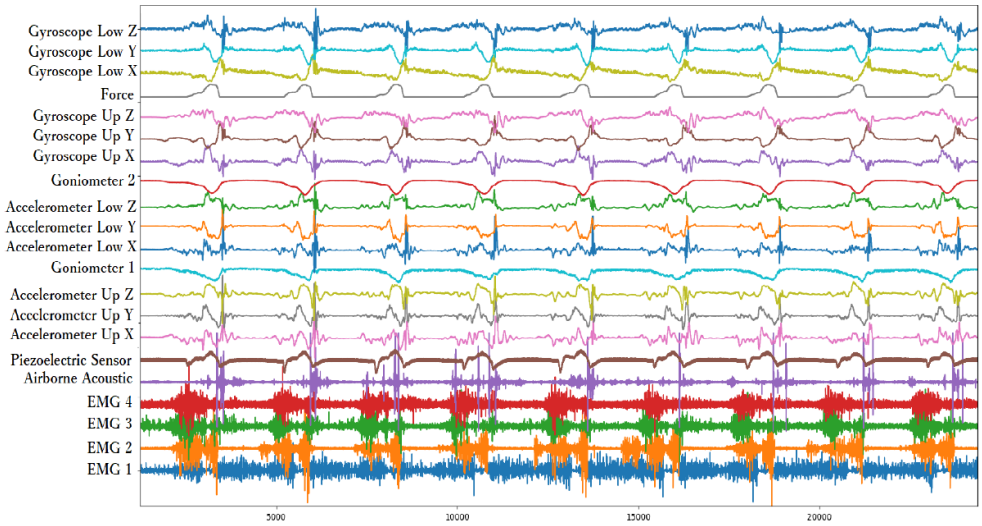


Figure 2.9 – An example of multi-sensorial data visualization.

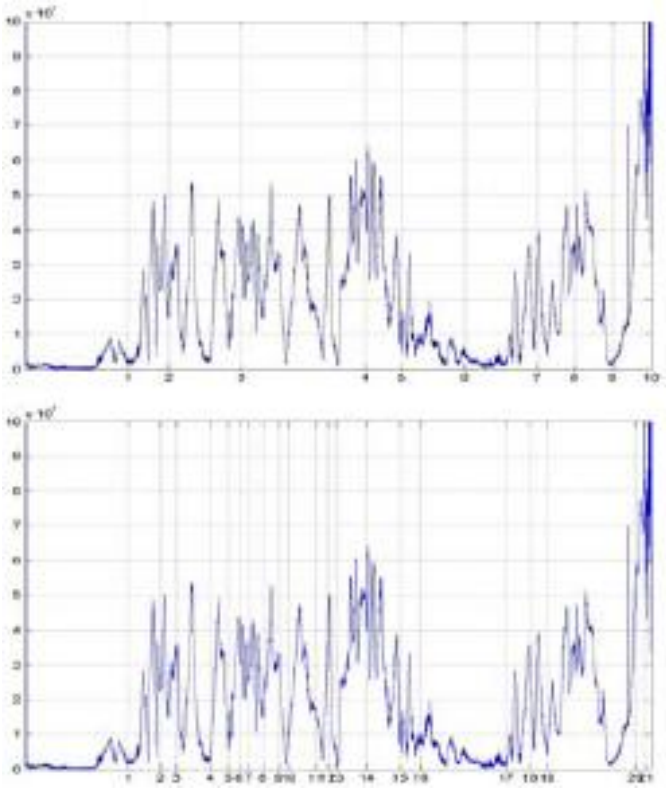


Figure 2.10 – Segmentation difference from subject to subject for the same task (vertical lines depict gesture boundaries). Top: the first annotator set 10 boundaries; bottom: the second annotator set 21 boundaries (Kahol et al., 2003).

The synchronization mechanism between video signals and biosignals will also affect the quality of the segmentation results — often, acoustic or optical signals are used to confirm the starting/ending synchronization. Besides, manual segmentation is more expensive in terms of time cost and labor cost.

In (Rebelo et al., 2013), simultaneous video recordings of the experiments were used to create references. In prior work (Palyafári, 2015), a student used in her master thesis the video from a head-mounted camera and an open-source annotation software *Advene* (Advene, 2021) to perform the later segmentation and annotation. Figure 2.11 demonstrates the interface of the *Advene* software. The video is played and stopped on the upper part, and the lower part shows all label segments. Through Figure 2.11, we can perceive the manual segmentation task's pressure and challenge.

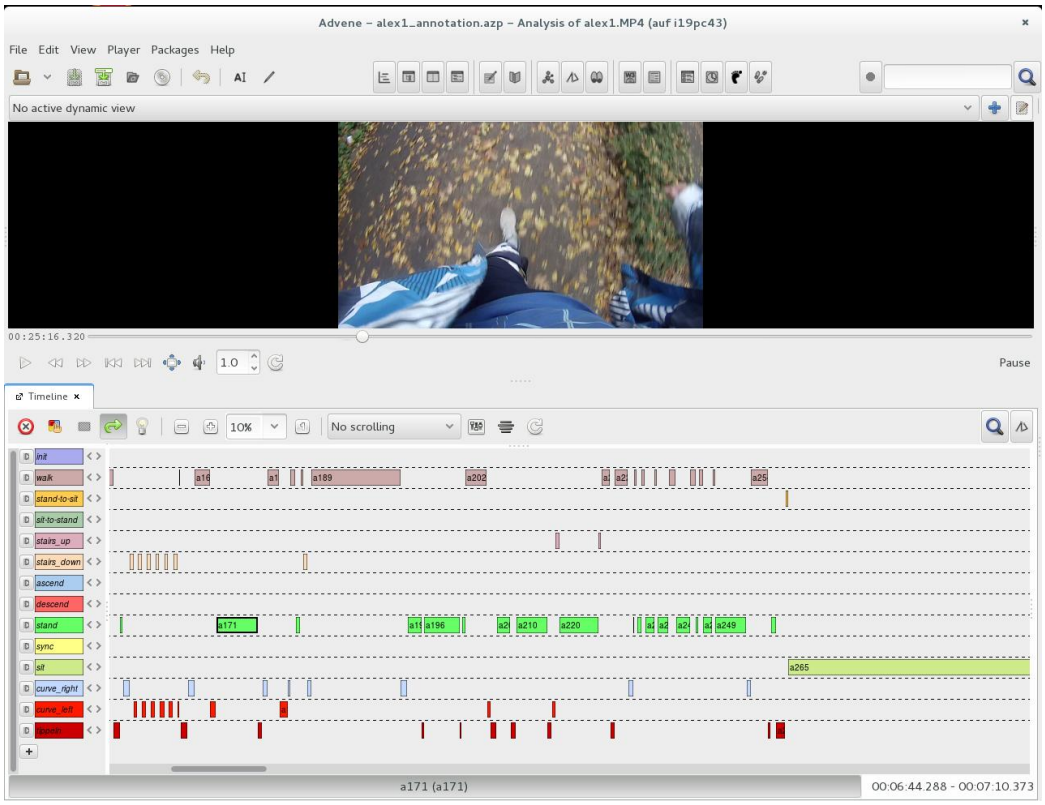


Figure 2.11 – Screenshot of the software *Advene* for manual segmentation and annotation (Palyafári, 2015).

Besides manual segmentation, modern machine learning methods, like *Gaussian Mixture Models* (GMMs), *Principal Component Analysis* (PCA) and *Probabilistic Principal Component Analysis* (PPCA) are being used to seg-

ment human activity automatically or semi-supervised (Barbič et al., 2004). (Guenterberg et al., 2009) applied signal energy to segment data. Moreover, appropriate features of long-term signals instead of segments can be extracted for the segmentation task, such as the research in (Ali and Aggarwal, 2001). In a bachelor thesis guided by us (Mai, 2019), *Convolutional Neural Networks* were applied to the multimodal biosignal dataset we collected for an automatic segmentation study. However, the experimental results were not satisfactory enough, and cannot directly be applied for follow-up research.

In our research, we applied the pushbutton of the *biosignalsplux Research Kit* (see Figure 2.4⑨) in our proposed semi-automated segmentation and annotation solution in order to save the research time of various automatic segmentation methods introduced above. In subsequent research, the applicability of the semi-automated segmented data has been verified for our research purpose during numerous experiments, so we have been applying this mechanism to our successively acquired datasets.

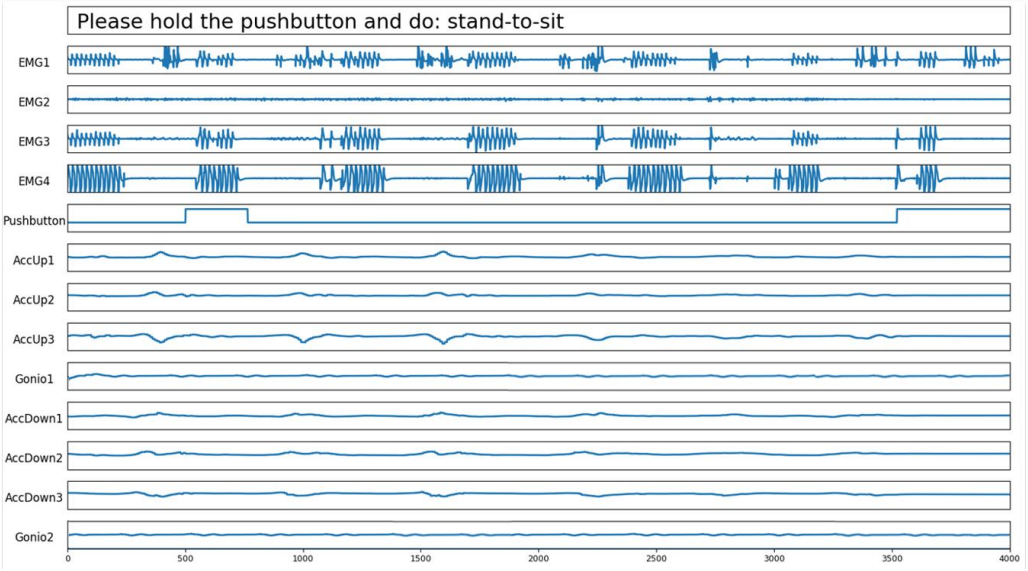


Figure 2.12 – Screenshot of the ASK software: the prompt shows the next activity to perform; the recorded biosignals over time are visualized for sanity check (Liu and Schultz, 2018).

The so-called protocol-for-pushbutton mechanism of segmentation and annotation has been implemented in the ASK software. When the "segmentation and annotation" mode is switched on during the data acquisition, a predefined activity sequence protocol will be loaded into the software, which prompts the user to perform the activities one after the other (see Figure 2.12). Each

activity is displayed on the screen one by one while the user controls the activity recording by pushing, holding, and releasing the pushbutton (Liu and Schultz, 2018). The user follows the instructions of the software step-by-step. For example, the user sees the instruction "Please hold the pushbutton and do: walk." The user prepares for it, then pushes the button and starts to do the activity "walk." She/he keeps holding the pushbutton while walking for a duration at will, then releases the pushbutton to finish this activity. With the release, the system displays the next instruction, e.g., "stand-to-sit," the process continues until the predefined acquisition protocol is fully processed.

The ASK software records all timestamps/sample numbers of each button push and button release during the data recording. These data are archived in CSV files as segmentation and annotation results for each activity. Since we synchronized all data at 1000 Hz, each sample represents data from 1 millisecond. As shown in Table 2.2, the first activity segment labeled "sit" lasts 2,517 samples, which corresponds to 2.517 seconds. The corresponding 2,517 samples form one segment for training the activity model "sit," or for the recognition evaluation.

Table 2.2 – An example of a CSV segmentation and annotation file (Liu and Schultz, 2018).

No.	Activity	From sample	To sample
1	sit	3647	6163
2	sit-to-stand	6901	9467
3	stand	11388	14181
4	stand-to-sit	16265	18882
5	sit	19396	22119
...	...	...	...

The time at the beginning of each recording and between the "release" and the "push," e.g., samples 0—3,646 or 6,164—6,900 in Table 2.2, correspond to the preparation time. Since preparation time is not part of the activity, we ignore the respective samples for training the models. However, these samples are still usable if we apply a shifting window on the data for continuous HAR.

Figure 2.13 depicts a data visualization fragment with a pushbutton channel. The protocol-for-pushbutton mechanism clearly separates two activities. We can also find that each activity contains three repeated motions. The activity between the two "hold-pushbutton"s is a "turn around," which is not included in the segmentation protocol (see Section 3.5.1).

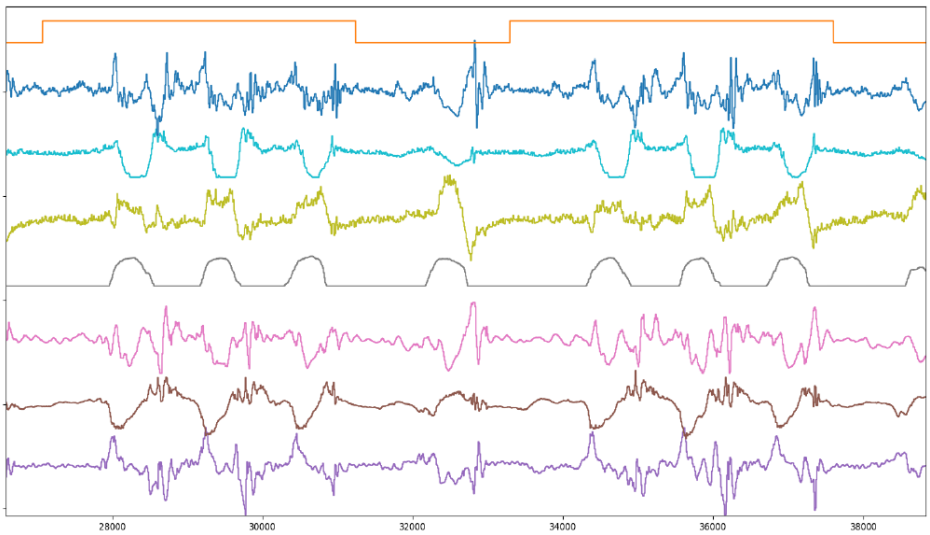


Figure 2.13 – Data visualization with the pushbutton channel on the top.

The protocol-for-pushbutton mechanism was implemented to reduce the time and labor costs of manual annotation. The resulting segmentations required little to no manual correction, and lay a good foundation for subsequent research. Nevertheless, this mechanism has some limitations:

- The mechanism can only be applied during acquisition and is incapable of segmenting archived data;
- Clear activity start-/endpoints need to be defined, which is impossible in cases like field studies;
- Activities requiring both hands are not possible due to participants holding the pushbutton;
- The pushbutton operation may consciously or subconsciously affect the activity execution;
- The participant forgetting to push or release the button results in subsequent segmentation errors;

None of these limitations, except forgetting to release the pushbutton, hold in a laboratory setting with clear instructions and protocols. Hence, misapplication of the pushbutton was addressed by real-time human monitoring of the incoming sensor signals, including the pushbutton channel, during acquisition. Additionally, a mobile phone video camera for post verification and adjustments was used (see Section 3.5.2).

In this dissertation, we also used an external open dataset, called *UniMiB SHAR* (see Section 3.6), to evaluate the modeling methods. This dataset's segmentation mechanism has a straightforward style: Each activity segment is three-second long, whether it is a "walk" or a "stand-to-sit." More details will be introduced in Section 3.6.

2.5 Biosignal Processing and Feature Extraction

HAR research is inextricably linked with signal processing. Compared to ordinary signals, biosignals have some unique properties that highlight the research topics of **bio**signal processing.

2.5.1 Definition and Characteristics of Biosignals

The meaning of "signal" can be described in several ways. Conceptually, the signal is a carrier of states or features from an object, such as a human body. From the mathematical point of view, the signal is the combination of several analog or digital functions or equations. In terms of physics, the signal is the objective substance or energy, including various forms of force, sound, heat, electricity, light, among others.

Biomedical signals, or biosignals, are carriers that carry the state or features of organisms. In Germany, biosignals are officially defined as **"Information that emanates from physical (possibly chemical) actions of the human body"**[1]. The word "biosignal" sometimes brings an intuitive notion of bioelectrical signals. In fact, biosignals in the academic sense include both electrical and non-electrical signals. Biosignals can be divided into several classes (Schultz, 2019):

- Acoustic biosignals. Examples: speech and sonar signals;
- Electrical biosignals. Examples: electrocorticogram (ECoG) and electroencephalogram (EEG) signals;
- Chemical biosignals. Examples: mass spectrum and chemoelectrical signals from body samples like blood, tears, or urine;
- Kinetic biosignals. Examples: inertial signals like linear acceleration and angular acceleration;

[1]Translated from the definition of DIN 44300, German Institute for Standardization.

- Optical biosignals. Examples: image sequence (video) and functional near-infrared spectroscopy signal (fNIRS);

- Thermal biosignals. Examples: temperature and thermal image signals.

Among the above-listed categories of biosignals, chemical and thermal biosignals cannot be applied to an HAR system so far because of the following two reasons: The precise measurements of these biosignals are usually invasive; In comparison of other types, these two types of biosignals sense and transmit the physical quantity with a relatively larger delay and high acquisition interval. The optical biosignal is mainly used in video-based HAR systems. This dissertation involves the remaining three types: kinetic, electrical, and acoustic biosignals because much research has shown their effectiveness for human activity recognition and analysis, as introduced in Section 2.2.

In practice, we found that biosignals have the following characteristics:

- Biosignals are relatively weak-amplitude signals. For example, the electrocardiogram (ECG) signal, usually considered having the strongest amplitude among bioelectrical biosignals, is only limited to the millivolt level (Saritha et al., 2008);

- Low-frequency biosignals often take place in the recording. For example, the EMG signal's frequency range is approximately between 5—500 Hz (Merletti and Di Torino, 1999). The characteristic of low frequency puts forward higher requirements for the signal amplifier's stability, and overcoming DC drift has become one of the key indicators of the biomedical devices;

- Interference or noise is particularly strong in biosignals. That is, the *Signal-to-Noise Ratio* (SNR) is low. In addition to the technical artifacts that can be effectively avoided, the physical interference of the surrounding environment and the biological interference of other signals emitted by the organism often significantly affect the primary signal's collection. For example, the main sources of artifacts during the ECG signal acquisition are the baseline wander mainly caused by respiration, and the high-frequency noise such as the EMG noise caused by the muscle activity (an ECG-EMG "crosstalk"). The more challenging point is that the interference signal often overlaps the target signal's frequency band, so the traditional filtering technology cannot be applied without losing information. For example, 50 Hz power frequency interference is in the effective frequency band of most biosignals.

These characteristics make it important to choose adequate and high qualitative equipment for research based on biosignals (see Section 2.2).

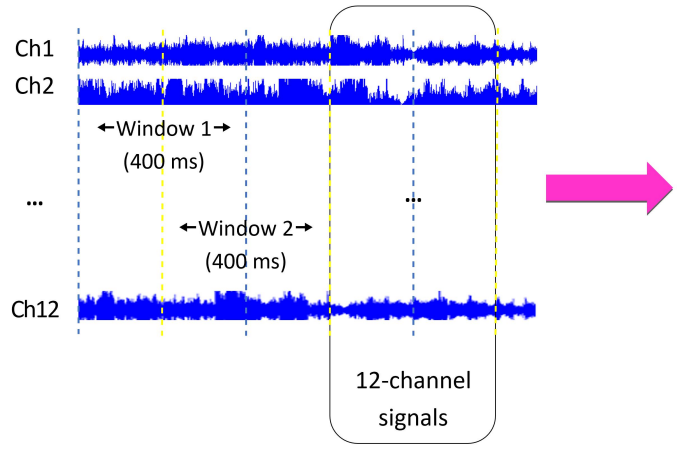
2.5.2 Digital Signal Processing and Feature Extraction

Some signal processing jobs can occur before segmentation (directly after or even during acquisition), such as filtering, amplification, noise reduction, and artifact removal. Another common example is normalization, which can also be applied to the whole collected biosignals instead of segments. Real-time systems need to use accumulated normalization because what we obtain from a real-time recording is always the continuous influx of short-term signal streams. Besides, as introduced in Section 2.4, feature extraction may also occur before segmentation, such as in the research of feature-based segmentation (Ali and Aggarwal, 2001).

In this dissertation, we do not study the signal processing that occurs before segmentation, but focus on segment-based *Digital Signal Processing* (DSP). Due to the characteristics of the biosignals introduced in section 2.5.1 and the demand for training and decoding, the biosignals captured by the sensors need to be preprocessed before further steps. The biosignals are firstly windowed using a specific window function with overlap. Second, a mean normalization is applied to the inertial and the EMG signals to reduce the impact of Earth acceleration and set the EMG signals' baseline to zero. Then, the EMG signals are rectified, a widely adopted signal processing method for muscle activities (Liu and Schultz, 2018).

Because multimodal biosignals for HAR systems are usually large-scale data, it is not common to use the raw data directly. Therefore, subsequently, features are extracted from each of the resulting windows.

Figure 2.14 illustrates with a schema the windowing and feature extraction on multichannel biosignals to build *Feature Vectors*. The 12-channel biosignals are windowed through a shifting window with a length of 400 ms and an overlap of 200 ms. Usually, the overlap between two adjacent windows can have a length chosen between 0 and the window length: the smaller the overlap length, the longer the training time of the model. Based on the windowing function, features will be extracted from each channel and form a *Feature Vector* of a window, which will be used for the follow-up tasks of training and recognition. The *Feature Vector* in the example of Figure 2.14 has a minimal dimension of twelve when only one feature is extracted from each signal channel.



Dim	12-Dimensional Feature Vector
1	$\overline{\sum_{i=0}^{N} x_{0i}}$: Acc upper X-axis
2	$\overline{\sum_{i=0}^{N} y_{0i}}$: Acc upper Y-Axis
3	$\overline{\sum_{i=0}^{N} z_{0i}}$: Acc upper Z-Axis
4	$\overline{\sum_{i=0}^{N} x_{1i}}$: Acc lower X-Axis
5	$\overline{\sum_{i=0}^{N} y_{1i}}$: Acc lower Y-Axis
6	$\overline{\sum_{i=0}^{N} z_{1i}}$: Acc lower Z-Axis
7	$\sqrt{\sum_{i=0}^{N} E_{0i}^2}$: EMG1
8	$\sqrt{\sum_{i=0}^{N} E_{1i}^2}$: EMG2
9	$\sqrt{\sum_{i=0}^{N} E_{2i}^2}$: EMG3
10	$\sqrt{\sum_{i=0}^{N} E_{3i}^2}$: EMG4
11	$\overline{\sum_{i=0}^{N} G_{0i}}$: E-Goniometer frontal
12	$\overline{\sum_{i=0}^{N} G_{1i}}$: E-Goniometer sagittal

Figure 2.14 – Example of building a *Feature Vector*: windowing and feature extraction for 400 ms window size.

Figure 2.14 implies two typical features, *average* and *Root Mean Square* (RMS), which have been applied in our biosignal processing work (Liu and Schultz, 2018), and also two of the mainly applied features in this dissertation's subsequent HAR research:

We denote the number of samples per window by N and the samples in the window by $(x_1, ..., x_N)$. For each window, the *average* feature is defined as:

$$avg = \frac{1}{N} \sum_{n=1}^{N} x_n. \tag{2.1}$$

For the EMG signal, we can, for instance, extract for each window the RMS feature:

$$RMS = \sqrt{\frac{1}{N} \sum_{k=1}^{N} x_n^2}. \tag{2.2}$$

The above-introduced two features are from the *statistical domain*. Besides, there are also various applicable features of time series in the *time domain*

and the *frequency domain*. (Figueira et al., 2016) summarized many features for HAR research in statistical, temporal, and spectral domains, as shown in Figure 2.15. Hence, numerous features can be extracted from various types of biosignals. In this dissertation's feature selection study, we applied over 30 features from the *Time Series Feature Extraction Library* (Barandas et al., 2020) jointly developed by research collaborators to select the best feature combination, which will be discussed in Section 5.4.

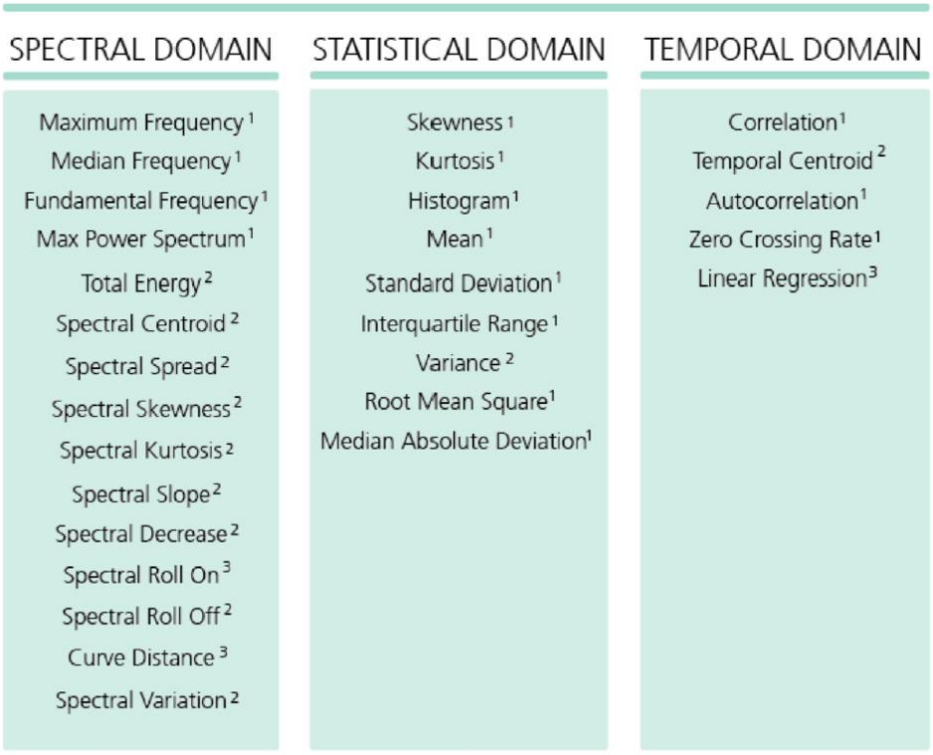


Figure 2.15 – Spectral, statistical and temporal domain features investigated in (Figueira et al., 2016). [1]Features already used in researches based on accelerometer signals; [2]features used in audio recognition (Peeters, 2004); [3]new features created and applied in (Figueira et al., 2016).

Features of different biosignals can be combined by early or late fusion, i.e., the *Feature Vectors* of single biosignal streams are either concatenated to form one multi-biosignal *Feature Vector* (*early fusion*), or recognition is performed on single biosignal *Feature Vectors*, and the combination is done on decision level (*late fusion*). In the HAR research in this dissertation, we rely on early fusion, which showed to outperform the late fusion strategy in the context of real-time HAR (Liu and Schultz, 2019).

2.6 Human Activity Modeling, Training, Recognition and Evaluation

Many machine learning methods for modeling have been applied to model human activities from sensor data effectively for later training and recognition, such as *Artificial Neural Networks* and *Hidden Markov Models*.

2.6.1 Human Activity Modeling Using Artificial Neural Networks

The field of *Artificial Neural Networks* (ANNs) tries to simulate and to create networks and devices inspired by neurobiology, to solve useful computational problems of the kind that biology does effortlessly (Hopfield, 1988). Thus, ANNs differ from conventional (digital or analog) computing machines that serve to replace, enhance or speed up human brain computation without regard to the organization of the computing elements and of their networking (Graupe, 2013). These networks are "neural" in the sense that they may have been inspired by neuroscience, but not because they are faithful models of biological neural or cognitive phenomena (Hassoun et al., 1995). ANNs are massively parallel computing systems consisting of an extremely large number of simple processors with many interconnections (Jain et al., 1996). It is necessary for the system to have a labeled directed graph structure where nodes perform some simple computations (Mehrotra et al., 1997). An ANN can also be described as mapping an input space to an output space, and this concept is analogous to that of a mathematical function (Priddy and Keller, 2005). From a signal processing perspective, ANNs' ability to adapt continuously to new data allows them to track changes in a signal over time, and their ability to learn from arbitrary, noisy data permits them to solve problems that cannot be handled adequately with some conventional statistical techniques (Abraham, 2005).

Deep Neural Networks (DNNs) (Miikkulainen et al., 2017) perform well in human activity recognition (Yang et al., 2015), (Oniga and Sütő, 2014), (Ordóñez and Roggen, 2016), (Kwon et al., 2015), (Wang et al., 2019). On this basis, many research works have shown the ability of *Convolutional Neural Networks* (CNNs) (Lee et al., 2017), (Ronaoo and Cho, 2015), (Ronao and Cho, 2016), (Zeng et al., 2014), (Ha et al., 2015) and *Recurrent Neural networks* (RNNs) (Inoue et al., 2018), (Deng et al., 2016), (Singh et al., 2017), (Murad and Pyun, 2017), (Arifoglu and Bouchachia, 2017) for HAR research.

Recently, *Residual Neural Network* (ResNet) models (He et al., 2016), which proved to be a compelling improvement of DNN for image processing, have also been used to research human activity recognition. A small amount of literature has already occurred in this direction (Tuncer et al., 2020), (Keshavarzian et al., 2019), (Long et al., 2019).

However, in many cases, researchers do not know each layer's specific physical meaning in neural networks. In contrast, the concept of "states" in the HMM definition-tuple (see Section 2.6.2) may have the better explanatory power of the activities' internal structure, which serves as the main modeling topology for this dissertation. In addition to the interpretability, HMM has other advantages for HAR study, such as the generalizability and reusability of models and states, and the analogizability between different research fields, which also laid the foundation for proposing and investigating our *Motion Units* activity modeling design (see Chapter 5).

2.6.2 Hidden Markov Model (HMM) and Continuous Density Hidden Markov Model

Markov Process (*Markov Chain*) (Ames, 1989) refers to the random process of transition from one state to another in the state-space. State-space describes a set of discrete states used in models. The notion of "state" can be thought of as the embodiment of the "memory" effect present in many physical systems, and such systems do not typically display arbitrary motions: the future behavior is usually determined by the past in some sense (Davis, 2002). The first-order *Markov chain* requires a "memoryless" nature: the probability distribution of the next state is only determined by the current state, where none of the events preceding it in the time series are related. This particular character of "memorylessness" is called *Markov Property* (Gurvits and Ledoux, 2005).

In the *Markov Model*, each state in the first-order *Markov chain* is directly visible to the observer. The essential parameters to determine are, therefore, the transitional probabilities between states. A *Hidden Markov Model* is a doubly stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols (Rabiner and Juang, 1986). The most efficient algorithms infer the implicit variables from the observations and then utilize these hidden variables for further applications, such as pattern recognition, where the most likely state sequence can be determined, and the model can be trained/optimized according to

the *Maximum Likelihood Estimation* (MLE). Each state has a probability distribution on the possible observations. In this sense, the output observation sequence can reveal information about the hidden state sequence to a substantial extent.

We denote the length of the observation sequence by T (i.e., $t = 1, 2, 3, ..., T$), the number of states S in the model by N, and the number of observation symbols by M. A discrete observation HMM is then formally defined as 5-tuple $\lambda = (S, V, \pi, A, B)$ (Rabiner, 1989):

- $S = \{s_1, s_2, ..., s_N\}$: the set of all possible states;
- $V = \{v_1, v_2, ..., v_M\}$: the discrete set of possible symbol observations;
- $\pi = \{\pi(s_i)\}, \pi(s_i) = P(q_1 = s_i \text{ at } t = 1)$: the initial state distribution at $t = 1$;
- $A = \{a_{ij}\}, a_{ij} = P(q_{t+1} = s_j | q_t = s_i), 1 \leq i, j \leq N$: state transition probability distribution, usually represented as a transition matrix of probabilities;
- $B = \{b_i(k)\}, b_i(k) = P(V_k \text{ at } t | q_t = s_i), 1 \leq i \leq N, 1 \leq k \leq M$: observation symbol probability distribution, usually represented as an emission matrix of probabilities.

In biosignal-based modeling research such as HAR and speech recognition, the observations are usually continuous space (e.g., features extracted from data segments) instead of a finite number of discrete symbols. Therefore, We need an extension to the 5-tuple basic HMM introduced above where observations $O_1, O_2, ..., O_T$ are continuous symbols, or more generally, continuous vectors (Rabiner and Juang, 1986), which cannot be simply described by a discrete set of possible symbol observations (V). For such a model, the $b_i(k)$ in the emission matrix B is replaced by a conditional distribution $b_j(O_t)$, called Probability Density Function (PDF), over the continuous observation space for each state (Nguyen, 2016):

$$b_j(O_t) = P_j(O_t | \theta_j), \tag{2.3}$$

where the PDF $P_j(O_t|\theta_j)$ can, theoretically, be any probability distribution, for example, an exponential distribution, but is usually restricted to be finite mixtures of simple parametric distributions, such as Gaussians (Rabiner and Juang, 1986). The notation θ_j denotes probabilistic parameters. For instance, if $P_j(O_t|\theta_j)$ is a normal distribution PDF, θ_j includes mean and variance values of Gaussians. This extended model is called *Continuous Density Hidden*

Markov Model (CDHMM), and all "HMM" mentioned in the following text of this dissertation refers to CDHMM.

There are three basic problems of interest that can be solved for the model to be useful in real-world applications (Rabiner, 1989):

- Problem 1: *Evaluation Problem*. Given the observation sequence $O = O_1, O_2, ..., O_T$ and a model λ, how do we efficiently compute $P(O|\lambda)$, the probability of the observation sequence, given the model?
- Problem 2: *Decoding Problem*. Given the observation sequence $O = O_1, O_2..., O_T$, and the model λ, how do we choose a corresponding state sequence $Q = q_1, q_2, ..., q_T$ which is optimal in some meaningful sense (i.e., best "explains" the observation O)?
- Problem 3: *Optimization Problem*. How do we adjust the parameters A, B, and π of the model λ to maximize $P(O|\lambda)$?

The research object of this dissertation is recognition, which can be placed as solving the second problem. An optimality criterion is almost trivially to find the single best path (state sequence) with the highest probability, representing the recognition procedure. A formal technique for finding this single best state sequence exists and is called the *Viterbi* algorithm (Rabiner and Juang, 1986), which can be directly applied using our in-house HMM-based toolkit for biosignals, *BioKIT* (see Section 2.6.4).

2.6.3 Sequence Modeling of Human Activity Using HMMs

HMMs are widely used for various activity recognition tasks, such as (Lukowicz et al., 2004) and (Amma et al., 2010). The former applies HMMs to an assembly and maintenance task, while the latter presents a wearable system that enables 3D handwriting recognition based on HMMs. In this so-called *Airwriting* system, the users write text in the air as if they were using an imaginary blackboard, while the handwriting gestures are captured wirelessly by accelerometers and gyroscopes attached to the back of the hand.

We use left-to-right HMM architectures of modeling as many applications of HMMs, such as speech recognition and online character recognition (Bishop, 2006), because its applicability and robustness for activity modeling have been verified by the preliminary modeling and recognition experiments in our laboratory (Palyafári, 2015) and performed gradually better in our follow-up research (see Chapters 4 and 5).

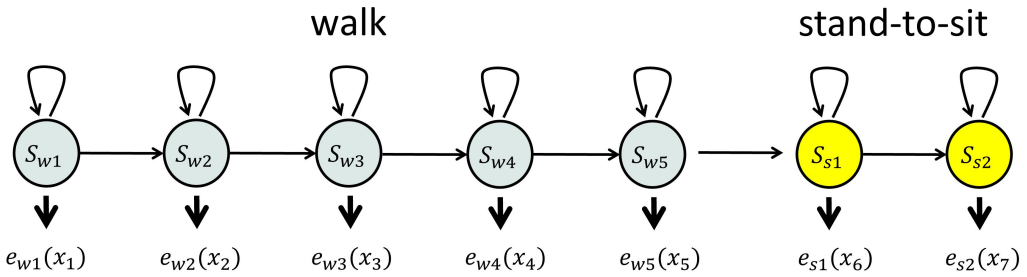


Figure 2.16 – A linear left-right HMM for an example of a typical human daily activity sequence "walk, then sit down," consisting of five-sate HMMs for the activity "walk" and two-state HMMs for the activity "stand-to-sit." S_i symbolize the states; the horizontal arrows show the transitions; e_i represent the distribution of the emission probabilities (PDF).

Figure 2.16 gives an example of modeling an activity sequence of two daily activities using HMMs. In this example, we assume that the activity "walk" has five (hidden) states, and the activity "stand-to-sit" has only two. By defining the essential components of the HMM modeling, the dictionary of the states for each activity, the transition probability distribution between the hidden states, the initial state distribution, and the emission probability distribution (PDF) from the hidden states to the observations, we can build comprehensive HMMs for all related activities in this work.

In our pilot study, we designed each activity with single state (see Section 5.1) or fixed number of states (see Section 5.2). In further research, we propose the concept of *Motion Units* based on phase and state partitioning, designing each activity with its individual adequate and reasonable number of states to improve the efficiency and correctness of modeling, training, and recognition. According to the experimental results, we found that many activity states can be generalized and applied in different activities, similar to speech recognition. Details will be described in Sections 5.5—5.7.

2.6.4 BioKIT: In-House HMM-Based Toolkit for Biosignals

BioKIT (Telaar et al., 2014) is an HMM-based toolkit to preprocess, model, and interpret biosignals such as speech, motion, muscle, and brain signals. This toolkit focuses on enabling researchers from various communities to pursue their experiments and integrate real-time biosignal interpretation into their applications.

BioKIT features a real-time capable token passing decoder with beam search. It makes beam search a greedy algorithm that only a predetermined number of the best partial solutions are kept as candidates. In order to streamline the beam search, the *beam prunings* in this toolkit eliminate all the *Viterbi*-paths that are not within a beam (i.e., a score difference) to the best element (Gehrig, 2015).

The terminology of *BioKIT* abstracts from the specific terminology to appeal to a wide audience in biosignal research (Telaar et al., 2014). The smallest meaningful unit in *BioKIT* is an **Atom**, representing a single HMM. A **Token** is a sequence of one or more Atoms, which compose *BioKIT*'s final recognition results. The **Dictionary** describes the mapping of Tokens to Atom sequences. The **Token Sequence Model** models occurrence probabilities of Token sequences. The **Feature Vector Scorer** computes emission probabilities of HMM states.

Table 2.3 relates *BioKIT* terms to common *Automatic Speech Recognition* (ASR) and HAR terminology.

Table 2.3 – The mapping of *BioKIT* terms to the ASR and the HAR terminology.

***BioKIT* Terms**	**ASR**	**HAR**
Token Sequence	Utterance	Human motion sequence (Gehrig, 2015), i.e., human activity in broad sense
Token	Word	Single human motion, which "human activity" refers to in this dissertation
Atom	Phone	State/sub-phase in single human motion

The definition of single human motion and human motion sequence in Table 2.3 is introduced in Section 1.1.

In the example illustrated in Figure 2.16, the Token Sequence consists of two Tokens, "walk" and "stand-to-sit," which contain five and two Atoms, separately.

We applied *BioKIT* directly in our research to model human activities, process biosignals, train HMM models using GMMs with *Split-and-Merge* algorithm (Zhang et al., 2003), (Li and Li, 2009), (Ueda et al., 2000), and build an end-to-end real-time recognition system.

2.6.5 Training, Recognition, Evaluation, and Iteration

The activity models for training and recognition can be built by any appropriate modeling method for HAR systems such as CNNs, RNNs, or HMMs which we focused on in this dissertation. As introduced in Section 2.6.3, in our research, each activity is modeled by an HMM, and a GMM represents the state emission probabilities. After training the model by taking the *Feature Vector* sequence from the "DSP & Feature Extraction" task (see Section 2.5.2) and the labels from the "Annotation" task (see Section 2.4), *BioKIT* follows the decoding of the activities (Tokens) based on the prepared *Feature Vector* sequence and provides the recognition results of the most probable activities. Top-N mode can be applied to generate N recognition results sorted by probabilities. In other words, the recognition result of the Top-N mode is not just one activity but N most probable activities.

A series of criteria and indicators will be applied to evaluate the prediction results using the ground truth provided by annotation: recognition accuracy, precision, recall, F-score, and confusion matrix, among others.

The results of the evaluation will contribute to improving the modeling, training, and feature selection. A so-called parameter tuning procedure alters the important parameters in training, such as the number of Gaussians for each emission model and the number of dimensions. New modification of the modeling, such as state amount and state generalization, can also happen during the re-training process. Iterative research on feature selection can also benefit from the evaluation task. Two loops in our HAR research pipeline 2.1.3 indicate two approaches of iterative improvement, guiding the research in Chapters 4 and 5, respectively:

- Feature study process (see Figure 2.17): "DSP & Feature Extraction $\rightarrow$ Modeling & Training $\rightarrow$ Recognition $\rightarrow$ Evaluation $\rightarrow$ DSP & Feature Extraction $\rightarrow$ Modeling & Training $\rightarrow \ldots \rightarrow$ Reporting";
- Parameter tuning or model optimization process (see Figure 2.18): "Modeling & Training $\rightarrow$ Recognition $\rightarrow$ Evaluation $\rightarrow$ Modeling & Training $\rightarrow \ldots \rightarrow$ Reporting."

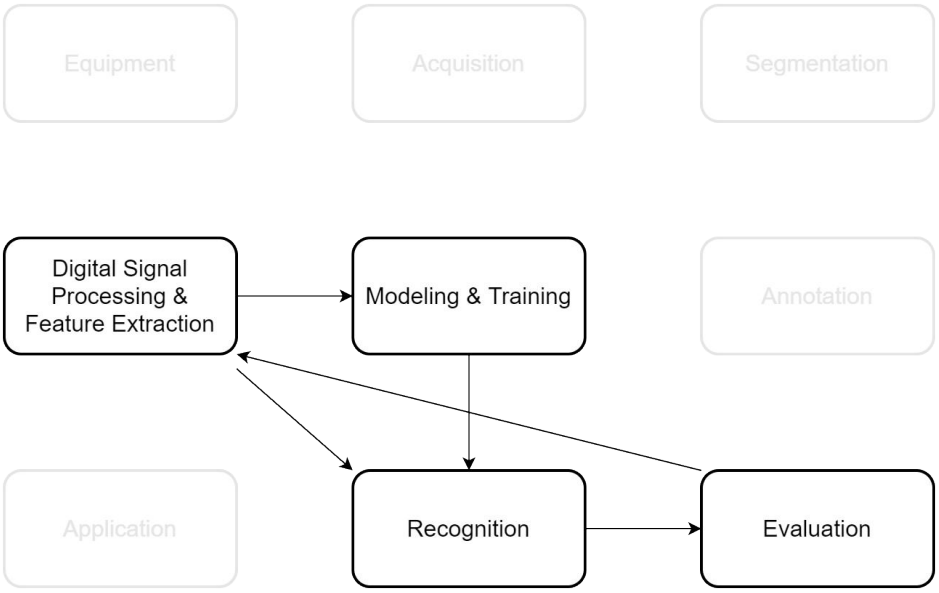


Figure 2.17 – The feature study process in the HAR research pipeline.

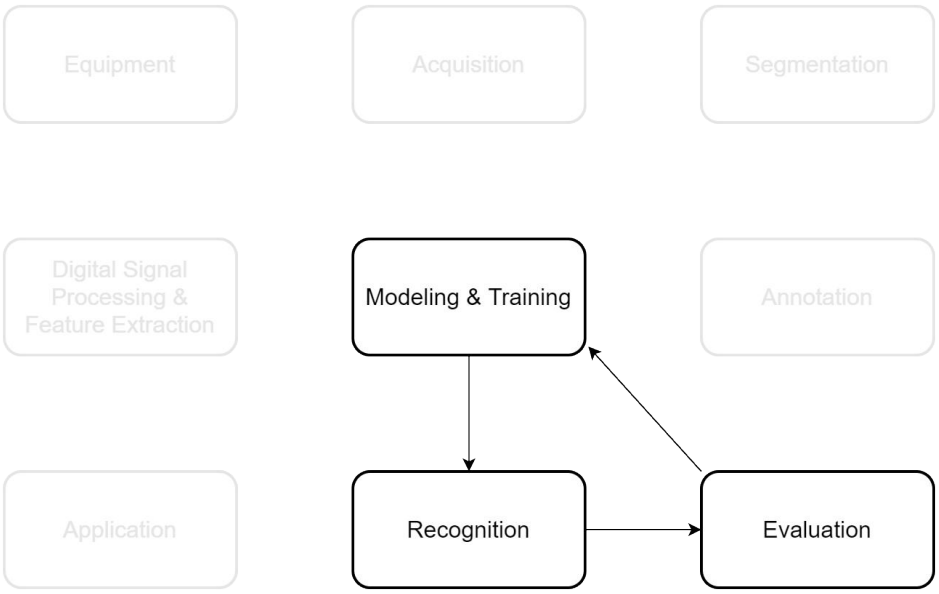
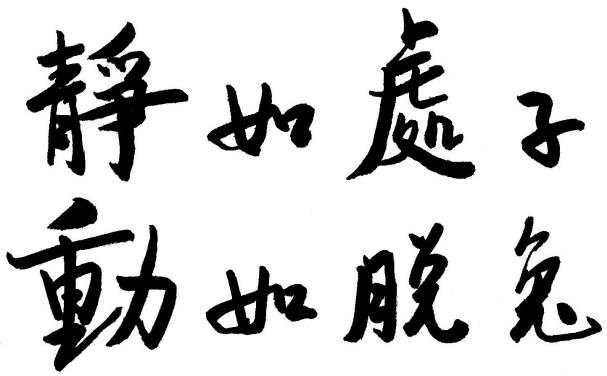


Figure 2.18 – The parameter tuning or modeling optimization process in the HAR research pipeline.

Chapter 3

Data Acquisition and Datasets



"*Be like a well-behaved noble girl when you need to be still;*

be like a lively field hare when you need to do actions."

Sun Tzu (544 BC — 496 BC), *The Art of War.*

HAR research relies on large amounts of data, which includes the laboratory data collections that meet in-house research goals, as well as the usage of external and public databases to verify models and methods. Therefore, data collection is an essential part of our entire HAR research work, for which we detail this extensive progress in this chapter.

3.1 Activities Included in the Datasets

It has been pointed out in Section 1.2 that we novelly use the knee bandage (see Figure 1.1) as a carrier of wearable sensors under the framework of the *Arthrokinemat* project (Arthrokinemat, 2021), making the collected datasets distinctive and more kinematically significant, and our HAR system focuses on assisting the early treatment of gonarthrosis. These factors determine which activities should be collected and studied in our research. As introduced in Section 2.1.2, the activity type studied in this dissertation is limited to ambulation activities, which is due to the following reasons:

- Most ambulation activities are whole-body (including locomotive and static) activities and emit distinguishable lower limb signals from a sensory perspective, which is very suitable for the research using sensor signals collected close to the knee;
- Compared with other activity type groups (transportation, phone usage, daily activities, fitness, military, and upper body), ambulation activities are more meaningful and irreplaceable for daily knee rehabilitation, which is decided in collaboration with kinesiologists of the *Institute of Sport and Sports Science* at *Karlsruhe Institute of Technology* (KIT) under the research framework of the *Arthrokinemat* project.

Specifically, which ambulation activities have been selected? The mainstay of early treatment of gonarthrosis is an adequate amount of proper movement. It results in muscular stabilization and fosters functional maintenance of the joints. In addition, movement is essential for the nutrition of both healthy and diseased cartilage (Liu and Schultz, 2018). Nevertheless, the knee joint with lesions should not be overloaded by these movements to not re-activate or further exacerbate gonarthrosis due to an inflammation of the joint, which would lead to even more pain for the patient and worsen the overall conditions. The goal is to technically assist the early treatment of gonarthrosis by discovering the right dose of daily movement, which affects the functionality of the joint positively while preventing movement-caused overload of the diseased joint (Liu and Schultz, 2019). Together with the

kinesiologists at KIT and based on the prior research (Palyafári, 2015), we selected the activities for our HAR research. The activities contained in each collected dataset will be given in their respective introductions in Sections 3.3—3.5. In these datasets' collections, the number of activities has increased gradually (see Table 3.1) due to the following reasons:

- The purpose of the data acquisition was gradually promoted: pilot dataset for proof-of-concept; advanced dataset for collaborative and person-independent functional research; comprehensive dataset for a research baseline;
- Based on the accumulated experimental results, we subdivided some activities into two activities. For example, some activities are subdivided into "left foot first" and "right foot first," resulting from applying only the right leg's knee bandage for sensor integration (see Section 3.5.1).

All the three in-house datasets are named as *"CSL{last two digits of the year} — number of **A**ctivities — number of **S**ignal-channels — number of **P**articipants"*, and abbreviated as *"CSL{last two digits of the year}"*. Table 3.1 concisely lists the statistics of all the datasets.

Table 3.1 – Statistics of the datasets applied in this dissertation. UMS: the *UniMiB SHAR* dataset; UMS9: the subset of the *UniMiB SHAR* dataset that includes nine activities of daily living; #: number of; l.: length.

Source	**In-house**			**External**	
Dataset	**CSL17**	**CSL18**	**CSL19**	**UMS**	**UMS9**
Sampling rate	1,000 Hz			50 Hz	
#Subjects	1	4	20	30	
#Channels	13	22	17	3	
#Activities	7	18	22	17	9
#Sequences	4	37	337		
#Segments	195	989	8,561	11,771	7,579
Total length	00:14:34	01:29:28	11:31:01		
Segmented l.	00:09:14	00:38:12	06:02:39	09:48:33	06:18:57
Raw data size	23.9 MB	225.3 MB	1.5 GB	81.4 MB	52.4 MB

In addition to the in-house datasets, Section 3.6 will briefly introduce an external dataset, *UniMiB SHAR* (see Table 3.1), serving to verify the applicability, replicability, and universality of our research results. Part of the research topics in this dissertation, such as feature space study, uses all the data of the *UniMiB SHAR* dataset (see Section 4.3), while other parts, such as *Motion Units* experiments (see Sections 5.6 and 5.7.3), only involves some

of the activities recorded in the *UniMiB SHAR* dataset, i.e., a subset of nine daily activities.

Before collecting data, we must first decide the sensors' position on the knee bandage.

3.2 Sensor Integration Scheme on the Knee Bandage

The usage of two triaxial accelerometers, four bipolar EMG sensors, and an electrogoniometer is proven to be effective for HAR in (Palyafári, 2015) and (Rebelo et al., 2013). On this basis, we added up to four additional sorts of sensors (12 channels), i.e., two triaxial gyroscopes, one airborne microphone, one piezoelectric microphone, and one force sensor, as candidates for sensor selection, as introduced in Section 2.2.5. We applied different combinations of sensors to collect three datasets successively from 2017 to 2019 based on the accumulative data acquisition experience and experimental results. However, their position and direction on the bandage, i.e., the sensors' integration scheme, which was also decided in collaboration with kinesiologists of the *Institute of Sport and Sports Science* at KIT based on their research experience in this field (Stetter et al., 2018), (Stetter et al., 2019a), (Stetter et al., 2019b), (Stetter et al., 2020) to capture ambulation activities, remained unchanged. Table 3.2 lists all measured muscles/body parts and sensor positions, and Figures 1.2 and 1.3 demonstrate intuitively the knee bandage integrated with all sensors and the position of the EMG electrodes, respectively.

The reasons for choosing the sensors in Table 3.2, including literature study, prior laboratory experiments, and references for sensor selection, have been clarified in 2.2. The number of EMG sensors (electrodes) and, consequently, their positioning depend mainly on the bandage's length and integration design, such as textile and hole-opening research. Generally, the muscle groups covered by a knee bandage include thigh and shin muscles. Two main groups of muscles are responsible for moving the knee: the extensor and the flexor muscles. The former muscle group on the frontal thigh is called "quadriceps," and the latter muscle group on the backside is called "hamstring." When the quadriceps contract, the knee extends, while the contraction of hamstring muscles makes the knee bend. Since not all major muscle groups of the muscles surrounding the knee can be well sensed in an integration scheme using knee bandage, one of the anterior thigh muscles (musculus vastus medialis), one of the posterior thigh muscles (musculus

biceps femoris), as well as two shin muscles (musculus tibialis anterior and musculus gastrocnemius) are selected.

Table 3.2 – Sensor placement and captured muscles/body parts.

Sensor	**Position/Muscle**
Accelerometer 1 (upper)	Thigh, proximal ventral
Accelerometer 2 (lower)	Shank, distal ventral
Gyroscope 1 (upper)	Thigh, proximal ventral
Gyroscope 2 (lower)	Shank, distal ventral
EMG 1 (upper-front)	Musculus vastus medialis
EMG 2 (lower-front)	Musculus tibialis anterior
EMG 3 (upper-back)	Musculus biceps femoris
EMG 4 (lower-back)	Musculus gastrocnemius
Electrogoniometer	Knee of the right leg, lateral
Airborne microphone	Knee of the right leg, lateral
Piezoelectric microphone	Knee of the right leg, lateral
Force Sensor	Between patella and bandage

3.3 Pilot Dataset CSL17 (CSL17-7A-12S-1P)

After finishing the ASK software development (see Section 2.3.1) and the first testing cycle, we applied it to collecting a pilot dataset *CSL17-7A-12S-1P* (CSL17) to validate the HAR research pipeline (see Section 2.1.3) and the software's practicability and robustness.

3.3.1 Sensors, Activities, and Acquisition Protocols

The CSL17 dataset consists of biosignals captured by **two triaxial accelerometers**, **four EMG sensors**, and **one biaxial electrogoniometer**, for a proof-of-concept study in the year of 2017. All sensors are placed on the bandage as described in Table 3.2. The difference between equipment applied in CSL17 data acquisition and Figure 1.2 is that CSL17 only uses two recording hubs. However, the overall appearance and integration style can still refer to this figure.

Seven basic daily activities are captured in this dataset: **"sit"**, **"stand"**, **"sit-to-stand"**, **"stand-to-sit"**, **"walk"**, **"walk-curve-left"**, and **"walk-curve-right"**. Five of these activities correspond to those described in (Rebelo et al., 2013), while the remaining two activities in this paper, "ascend"

and "descend," were replaced by "walk-curve-left" and "walk-curve-right" in our data acquisition. "Walk-curve-left"/"walk-curve-right" means left/right turn with an arbitrary angle using several walking steps, while the meaning of the other five activities can be intuitively known from their label names. All seven activities are the most common ambulation activities, which appear in many other public datasets and are widely researched for HAR using various sensors and positioning designs.

In order to make data acquisition more efficient, we tried to assign these seven activities to different recording protocols. It is conceivable that using a chair, "sit," "sit-to-stand," "stand," and "stand-to-sit" can be collected consecutively and repeatedly (of course, due to the protocol-for-pushbutton mechanism, there must be a pause between two consecutive activities for the pushbutton operation), but activities like "walk" obviously cannot be seamlessly inserted into this sequence. As a result, the seven activities were organized in two clusters based on their logical, sequential correlation — *"stay-in-place"* and *"move-around"*, which results in two acquisition protocols of activity sequence as follows (<push+hold> and <release> indicate the operation of the pushbutton under the "segmentation and annotation" mode introduced in Section 2.4):

- **Protocol 1** *"stay-in-place"*:
  <push+hold> sit <release> → <push+hold> sit-to-stand <release> → <push+hold> stand <release> → <push+hold> stand-to-sit <release> →...
- **Protocol 2** *"move-around"*:
  <push+hold> walk any number of steps (similarly hereinafter) <release> → <push+hold> turn left with an arbitrary angle using any number of steps <release> → <push+hold> walk <release> ⇀ (turn around in place) ⇀ <push+hold> walk <release> → <push+hold> turn right with an arbitrary angle using any number of steps <release> → <push+hold> walk <release> → (turn around in place) →...

3.3.2 Statistics and Analysis

Four recording sessions (two for each acquisition protocol) from one male subject were collected on the same day. For a proof-of-concept study, we wanted to record about 15 minutes of data. Since the activities in "stay-in-place" sessions have a shorter duration, 25 repetitions were required. In contrast, the "move-around" session only requires 17 repetitions, so that the total recording length of each session is about the same. Only one "move-

around" session is about 100 seconds longer than the others because we intended to record some longer "walk," "walk-curve-left," and "walk-curve-right" activities to provide data diversity for primitively verifying the training and recognition robustness. Table 3.3 summarizes the recording duration of all sessions.

Table 3.3 – Duration of each session in the CSL17 dataset.

Session	Acquisition Protocol	Total length
1	"stay-in-place"	00:03:10
2	"stay-in-place"	00:03:18
3	"move-around"	00:02:58
4	"move-around"	00:05:09
Sum	4 sessions	00:14:34

Due to the feasible organization of the acquisition protocols, recording sessions were executed very efficiently with only small amounts of not-planned-activity frames. The four recording sessions' total length adds up to about 15 minutes, of which 9.23 minutes of data have been segmented and annotated. We are aware that this dataset is small-scaled and neither sufficiently nor necessarily large enough to establish reliable recognition accuracy for daily activities. However, the purpose of this pilot dataset was to verify the training and recognition functionality of the ASK software, and to rapidly prototype an end-to-end wearable real-time human recognition system,

The "segmentation and annotation" mode (see Section 2.4) of the ASK software was used in the data collection to segment the recorded data in real-time and to generate for each segment the corresponding labels. In this mode, the ASK software can accumulate recording statistics such as the number of activity segments, the total effective length over all segments, and the minimal/maximal/mean length of each activity (see Table 3.4).

For the statistics in Table 3.4, the following points need to be pointed out:

- The number of activity segments of five activities is smaller than the preset number of repetitions, which is because, in post verification, we removed some wrong segments due to the misoperation of the pushbutton.
- The "walk" activity's number of segments and total duration are considerably larger than other activities because it appears more times in the acquisition protocol, and it is also a relatively longer activity natively;

Table 3.4 – Statistics of the segmented corpus in the CSL17 dataset. The minimum (Min.), maximum (Max.), mean, and standard deviation (std.) values are in seconds. #Seg.: number of segments.

Activity	**Min.**	**Max.**	**Mean±std.**	**#Seg.**	**Total**
sit	1.637	7.777	3.20 ± 1.35	25	00:01:20
stand	1.491	17.818	3.54 ± 3.15	23	00:01:21
sit-to-stand	1.444	2.566	1.97 ± 0.30	24	00:00:47
stand-to-sit	1.189	2.836	1.92 ± 0.47	23	00:00:44
walk	1.351	4.566	2.58 ± 0.62	67	00:02:53
walk-curve-left	1.811	12.997	3.64 ± 3.09	17	00:01:02
walk-curve-right	1.563	18.117	4.14 ± 4.74	16	00:01:06
Total	1.189	18.117	2.84 ± 2.18	195	00:09:14

- Most segments of the "sit-to-stand" and "stand-to-sit" activities, which can be colloquially called "stand up" and "sit down," are inherently shorter than other activities' segments;

- The maximum segment length of the "walk-curve-left" and "walk-curve-right" activities is unusual but under control. The subject walked in a loop sometimes in order to produce data of certain special situations for testing the stability of the training and recognition, which also leads to the longer duration of the recording session 4.

The original intention of the CSL17 protocol design and repetition presetting was to balance each activity's total length. However, the result did not turn out as expected. We learned this lesson from the data collection of CSL17 and subsequent CSL18, and for CSL19, we gradually improved the protocol design, which aims to have all activities reach well-balanced in the number of segments, making more confidence in the modeling, training, and recognition research.

From Table 3.4, we can also see that, no activity in the CSL17 dataset is shorter than 1.189 seconds. This a priori information helps us decide the initial values of some training parameters beforehand, such as window length and overlap length.

We used the CSL17 dataset to verify the HAR research pipeline's reasonableness (see Section 2.1.3) and the integrity and correctness of the data recorded by ASK Software (see Section 5.1.2). Subsequently, the up-to-date real-time HAR system ASKED (see Section 5.1.3) applied the CSL17 dataset as the

fundamental training dataset for the investigation of the end-to-end system's performance.

3.4 Advanced Dataset CSL18 (CSL18-18A-21S-4P)

After ensuring that the data collection function of the ASK Software runs efficiently without errors and obstacles, we continued to record a larger dataset *CSL18-18A-21S-4P* (CSL18) of 18 activities using the ASK software.

3.4.1 Sensors and Activities

The recording process was executed in cooperation with research partners from the *Institute of Sport and Sports Science*, KIT, in their laboratory. Besides the increase in the number of activities, we also extended the types of sensors by four additional ones, i.e., **one airborne microphone**, **one piezoelectric microphone**, **two triaxial gyroscopes**, and **one force sensor**, which increases the total recording channels to 21. The sensors' positioning on the bandage is consistent with the description in Table 3.2, and Figure 1.2 illustrates the knee bandage integrated with all sensors.

The five basic activities, i.e., **"sit"**, **"stand"**, **"sit-to-stand"**, **"stand-to-sit"**, and **"walk"** in the CSL17 dataset remain in the CSL18 dataset, while **"Walk-curve-left (90°)"** and **"walk-curve-right (90°)"** in the CSL18 dataset are more restrictive than the two in the CSL17 dataset in the sense of the turning angle: Each "walk-curve" activity undergoes a 90° overall turn. The other 11 additional activities are: **"walk-upstairs"**, **"walk-downstairs"**, **"spin-left"**, **"spin-right"**, **"run"**, **"V-cut-left"**, **"V-cut-right"**, **"shuffle-left"**, **"shuffle-right"**, **"jump-one-leg"**, and **"jump-two-leg"**. Most of the activities are self-explanatory or have been introduced in Section 3.3.1. The remaining are defined as follows:

- Spin-left and spin-right: can be described as the "Left face!" or "Right face!" action in the army (but in daily situations, not so stressful as in military training);
- Run: a type of gait characterized by an aerial phase in which all feet are above the ground (Rubenson et al., 2004), forward, with several steps;
- V-cut-left and V-cut-right: a direction change with an acute angle during running;

- Shuffle-left and shuffle-right: a type of gait characterized by an aerial phase in which all feet are above the ground (Rubenson et al., 2004), leftward or rightward, starting with the left/right foot, the other following, with several steps;

- Jump-one-leg and jump-two-legs: jumping up vertically in place using the bandaged leg/both legs;

The CSL18 data collection work is carried out in cooperation with our research partner's laboratory. Each participant first took part in the Vicon System (Vicon, 2021) motion capture recording under the instruction of our research partner, and then wore the bandages integrated with sensors we provided to participate in the CSL18 data acquisition. Therefore, during the recording sessions, we required the participants to perform the activities following the same recording process as they did in the motion capture recording. We did not specify the number of steps, mainly for the "walk," "run," "walk-curve," and "V-cut" activities, due to a compromise between the two parties' data on collaborative project research. For the subsequent research on activity modeling, we have strictly limited the number of steps of each activity in the acquisition protocols of CSL19. Therefore, the unlimited steps of CSL18's gait-based activities can serve as a suitable material for the continuous HAR that we are about to research.

3.4.2 Statistics and Analysis

We originally recorded data from seven male subjects. However, after we conducted some early data application work, such as training and recognition experiments (see Section 5.2.2), we had to drop three of the seven subjects' recordings due to technical issues, including frame drops during the recording, resulting in the remaining 1.5 hours of data from four male subjects, of which 38 minutes have been segmented and annotated by the protocol-and-pushbutton mechanism (see Section 2.4). Table 3.5 gives the number of activity segments, the total effective length over all segments, and the minimal/maximal/mean length of the 18 activities.

An imbalance of occurrences can be observed for "run" and "walk" and is explained by their repeated use in different acquisition protocols applied similarly to the description in Section 3.3.1. This imbalance is welcomed as it reflects expectations in uncontrolled settings and allows for a more detailed model better discriminating similar activities (Hartmann et al., 2020).

Table 3.5 – Statistics of the segmented corpus in the CSL18 dataset. The minimum (Min.), maximum (Max.), mean, and standard deviation (std.) values are in seconds. #Seg.: number of segments.

Activity	**Min.**	**Max.**	**Mean±std.**	**#Seg.**	**Total**
sit	1.329	5.089	3.11 ± 0.91	38	00:01:58
stand	1.759	5.129	3.11 ± 0.77	37	00:01:55
sit-to-stand	0.939	3.589	1.73 ± 0.55	42	00:01:13
stand-to-sit	1.029	3.449	1.78 ± 0.57	42	00:01:15
walk	1.579	5.179	2.93 ± 0.69	198	00:09:41
walk-curve-left (90°)	1.759	4.269	3.01 ± 0.59	44	00:02:12
walk-curve-right (90°)	1.649	3.789	2.76 ± 0.51	45	00:02:04
walk-upstairs	1.989	5.159	3.93 ± 0.72	43	00:02:49
walk-downstairs	1.769	5.259	3.61 ± 0.81	45	00:02:42
spin-left	0.725	3.249	1.41 ± 0.57	54	00:01:16
spin-right	0.666	2.279	1.26 ± 0.38	47	00:00:59
run	1.179	3.139	1.98 ± 0.53	86	00:02:50
V-cut-left	0.709	2.209	1.13 ± 0.35	43	00:00:49
V-cut-right	0.679	1.699	1.14 ± 0.29	37	00:00:42
shuffle-left	1.129	4.089	2.23 ± 0.86	47	00:01:45
shuffle-right	0.969	4.379	2.18 ± 0.93	47	00:01:43
jump-one-leg	0.749	2.159	1.51 ± 0.29	47	00:01:11
jump-two-leg	0.999	1.969	1.45 ± 0.23	47	00:01:08
Total	0.666	5.258	2.32 ± 1.03	989	00:38:12

3.5 Comprehensive Dataset CSL19 (CSL19-22A-17S-20P)

The *CSL19-22A-17S-20P* (CSL19) dataset is a follow-up to the CSL18 dataset and was recorded at the *Cognitive Systems Lab* (CSL) in a controlled laboratory environment.

3.5.1 Sensors, Activities, and Acquisition Protocols

In contrast to the CSL18 dataset, the two microphones, one goniometer, and one force sensor channel have been removed since the sensor selection experiments indicated that these sensors do not provide additional relevant information (see Sections 4.1.3 and 4.2.1). Therefore, the CSL19 dataset has 17 channels of biosignals, compared to 21 channels of the CSL18 datasets.

The 22 activities in the CSL19 dataset are comprised of ambulation activities, which are decided, as introduced in Section 3.4.1, based on the consensus reached with the research partners under the framework of the *Arthrokinemat* project (see Section 1.2). Compared to the CSL18 dataset, the four additional activities in the CSL19 dataset are the subdivision from four original activities in the CSL18 dataset, as the clues found from CSL18 data visualization and analysis indicated. For example, "spin-left" in the CSL18 dataset is divided into *"spin-left-left-first"* and *"spin-left-right-first"*, denoting which foot should be moved first. Similarly, "spin-right," "V-cut-left," and "V-cut-right" are also divided into two activities in regard to the first-moved foot. The activities mentioned above are subdivided because they only involve one gait, and we only use the sensors placed on the right-leg bandage. Therefore, the "left foot first" and "right foot first" of these activities will lead to very different activity modeling results. On the contrary, for activities involving multiple gaits/steps, such as "walk" (and its derivative activities), "run," and "shuffle"s, we did not further subdivide them. Instead, we restricted in the protocols the number of gaits for each segment of these activities to three and defined the left foot as the start. Therefore, it is of little significance to subdivide them into the "left foot first" or "right foot first," since the activity modeling research of these activities focuses on the gait (see Section 5.3.2).

In contrast to the recording events of the CSL17 and the CSL18 datasets, the acquisition protocols of the CSL19 dataset recording events have been strictly and normatively designed:

- For the subsequent study of activity modeling, we need to restrict the core parameters in each activity segment, the most important being the number of gaits and the angle of turning;
- As introduced in Sections 3.3.2 and 3.4.2, in the CSL17 and the CSL18 datasets, each activity's number of segments did not turn out as expected balanced, which is not ideal for training and recognition research. In order to solve this problem, most of the CSL19 acquisition protocols contain only one activity. However, there are two protocols with two activities and one protocol with four activities because these activities can be practically and logically recorded one after another in a sequence, which also keeps the balance of the activity occurrences. To follow the logical sequence of the activities and the protocol-for-pushbutton mechanism, the order of the activities in these three multi-activity protocols must be observed during recording.

The 22 activities and the 17 acquisition protocols are described as follows:

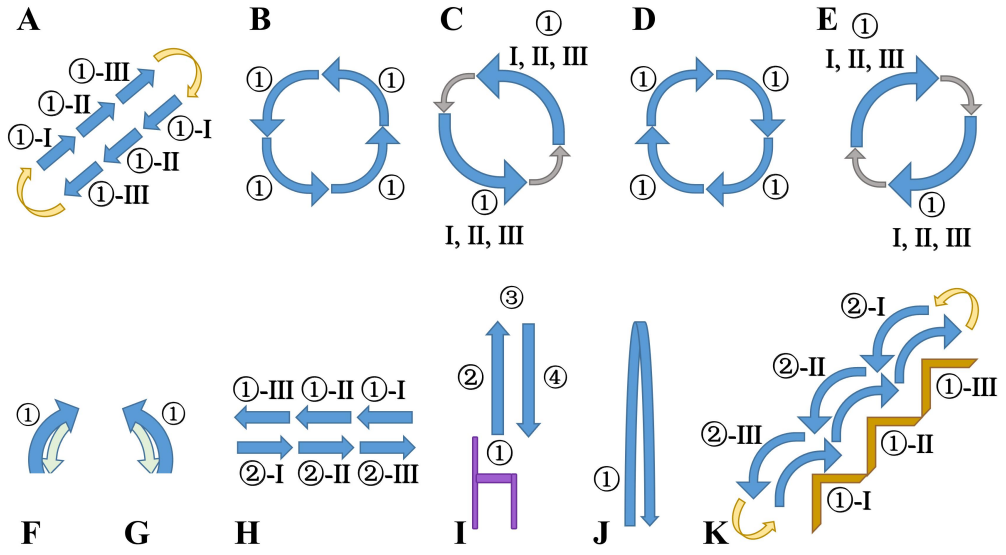


Figure 3.1 – Diagrammatic sketch of the recording protocols. Blue arrows: activities or gaits; I, II, III: gait cycles; yellow arrows: turn around (180°) in place; gray arrows: turn left/right (90°) in place; green arrows: return backward to the start point; purple: a chair; brown: stairs. Table 3.6 describes the perspective and the corresponding protocols of each sub-diagram in detail.

- Protocol 1: **walk** (Figure 3.1(A))
  <push+hold> walk forward with three gait cycles, left foot starts, i.e., left-right-left-right-left-right <release> → (turn around in place) →...
  20 repetitions (20 activities with 60 gait cycles) per subject.
  Note: the "turn around in place" between two "walk"/"run" activities is due to the limited space in our laboratory.

- Protocol 2: **walk-curve-left (90°)** (Figure 3.1(B))
  <push+hold> turn left 90° with three gait cycles, left foot starts <release> → (turn left 90° in place) →...
  20 repetitions (20 activities with 60 gait cycles) per subject.

- Protocol 3: **spin-left-left-first** (Figure 3.1(C))
  <push+hold> turn left 90° in one step, left foot starts <release> →...
  20 repetitions (20 activities with 20 gait cycles) per subject.

- Protocol 4: **spin-left-right-first** (Figure 3.1(C))
  <push+hold> turn left 90° in one step, right foot starts <release> →...
  20 repetitions (20 activities with 20 gait cycles) per subject.

Table 3.6 – Detailed description of the sub-diagrams in Figure 3.1: the perspectives and the corresponding protocols with activities.

Sub-diagram	Perspective	Protocol	Activity
Figure 3.1(A)	top view	1	① walk
		8	① run
Figure 3.1(B)	top view	2	① walk-curve-left (90°)
Figure 3.1(C)	top view	3	① spin-left-left-first
		4	① spin-left-right-first
Figure 3.1(D)	top view	5	① walk-curve-right (90°)
Figure 3.1(E)	top view	6	① spin-right-left-first
		7	① spin-right-right-first
Figure 3.1(F)	top view	9	① V-cut-left-left-first
		10	① V-cut-left-right-first
Figure 3.1(G)	top view	11	① V-cut-right-left-first
		12	① V-cut-right-right-first
Figure 3.1(H)	front view	13	① shuffle-left
			② shuffle-right
Figure 3.1(I)	side view	14	① sit
			② sit-to-stand
			③ stand
			④ stand-to-sit
Figure 3.1(J)	side view	15	① jump-two-leg
		16	① jump-one-leg
Figure 3.1(K)	side view	17	① walk-upstairs
			② walk-downstairs

- Protocol 5: **walk-curve-right (90°)** (Figure 3.1(D))
  <push+hold> turn right 90° with three gait cycles, left foot starts <release> → (turn right 90° in place) →...
  20 repetitions (20 activities with 60 gait cycles) per subject.

- Protocol 6: **spin-right-left-first** (Figure 3.1(E))
  <push+hold> turn right 90° in one step, left foot starts <release> →...
  20 repetitions (20 activities with 20 gait cycles) per subject.

- Protocol 7: **spin-right-right-first** (Figure 3.1(E))
  <push+hold> turn right 90° in one step, right foot starts <release>→...
  20 repetitions (20 activities with 20 gait cycles) per subject.

- Protocol 8: **run** (Figure 3.1(A))
  <push+hold> go forward at fast speed with three gait cycles, left foot starts <release> → (turn around in place) →...
  20 repetitions (20 activities with 60 gait cycles) per subject.
- Protocol 9: **V-cut-left-left-first (30°)** (Figure 3.1(F))
  <push+hold> turn 30° left forward in one step at jogging speed, left foot starts <release> → (return backward to the start point with three steps, left foot starts) →...
  20 repetitions (20 activities with 20 gait cycles) per subject.
- Protocol 10: **V-cut-left-right-first (30°)** (Figure 3.1(F))
  <push+hold> turn 30° left forward in one step at jogging speed, right foot starts <release> → (return backward to the start point with three steps, left foot starts) →...
  20 repetitions (20 activities with 20 gait cycles) per subject.
- Protocol 11: **V-cut-right-left-first (30°)** (Figure 3.1(G))
  <push+hold> turn 30° right forward in one step at jogging speed, left foot starts <release> → (return backward to the start point with three steps, left foot starts) →...
  20 repetitions (20 activities with 20 gait cycles) per subject.
- Protocol 12: **V-cut-right-right-first (30°)** (Figure 3.1(G))
  <push+hold> turn 30° right forward in one step at jogging speed, right foot starts <release> → (return backward to the start point with three steps, left foot starts) →...
  20 repetitions (20 activities with 20 gait cycles) per subject.
- Protocol 13: **shuffle-left** + **shuffle-right** (Figure 3.1(H))
  <push+hold> move leftward with three lateral gaits cycles, left foot starts, right foot follows <release> → <push+hold> move rightward with three lateral gaits cycles, right foot starts, left foot follows <release> →...
  20 repetitions (40 activities with 120 gait cycles) per subject.
- Protocol 14: **sit** + **sit-to-stand** + **stand** + **stand-to-sit** (Figure 3.1(I))
  <push+hold> sit <release> → <push+hold> stand up <release> → <push+hold> stand <release> → <push+hold> sit down <release>→...
  20 repetitions (80 activities) per subject.

- Protocol 15: **jump-one-leg** (Figure 3.1(J))
  <push+hold> squat, jump upwards using the bandaged right leg, land in <release> →...
  20 repetitions (20 activities) per subject.

- Protocol 16: **jump-two-leg** (Figure 3.1(J))
  <push+hold> squat, jump upwards using both legs, land in <release> →...
  20 repetitions (20 activities) per subject.

- Protocol 17: **walk-upstairs** + **walk-downstairs** (Figure 3.1(K))
  <push+hold> go up six stairs with three gait cycles, left foot starts <release> → (turn around in place) → <push+hold> go down six stairs with three gait cycles, left foot starts <release> → (turn around in place) →...
  20 repetitions (40 activities with 120 gait cycles) per subject.

In each participant's recording sessions, each protocol was executed one by one in order, and the order was not disrupted in most cases. It should be pointed out that the order of the protocols does not affect the collected data. In one of the 20 recording sessions, we had to change the protocol order during the data acquisition to temporarily vacate part of the laboratory room due to sporadic sharing issues.

To some extent, the sequential execution method loses the randomness of human activity but ensures the efficiency, usability, and data quality of laboratory collection as much as possible, which has been verified in the subsequent experiments, under limited time and human resources.

The number of repetitions/to-record activities per each protocol is a pre-designed plan. In the post verification, a few non-conformity and erroneous segments were removed.

3.5.2 Post Verification of the Segmentation Mechanism

Although the "segmentation and annotation" mode of the ASK software was switched on to segment and annotate the recorded data efficiently, a mobile phone video camera was used in addition to record the whole biosignal acquisition sessions to manually correct the human misoperation of pushing/holding/releasing after the data recording.

After each recording event with one subject, the collected data and the automatically generated segments with annotation labels were examined thoroughly based on the video. Segments with minor human-factor errors were corrected by shifting the start-/endpoints forward/backward a short distance manually, while segments with problems that cannot be easily corrected were discarded, which leads to the divergence between the activity occurrences in Table 3.7. A script to automatically detect the activity length outlier was also implemented to assist the segmentation verification. After finishing the correction and verification, we deleted all recorded videos to preserve privacy.

3.5.3 Statistics and Analysis

The CSL19 dataset contains 11.52 hours of data (of which 6.03 hours have been segmented and annotated) from 20 subjects, 5 female and 15 male. Table 3.7 gives the number of activity segments, the total effective length over all segments, and the minimal/maximal/mean length of the 22 activities.

By analyzing the duration distribution of each activity of all subjects in histograms, we find that all activities' duration over all segments approximately accords with the normal distribution. The distribution of the activities "sit" and "stand" deviates slightly, as they can last arbitrarily long. Figure 3.2 illustrates some examples of activity duration histograms. The area under the curve equals the total number of segment occurrences within 200-millisecond intervals. Appendix A provides all activities' duration histograms.

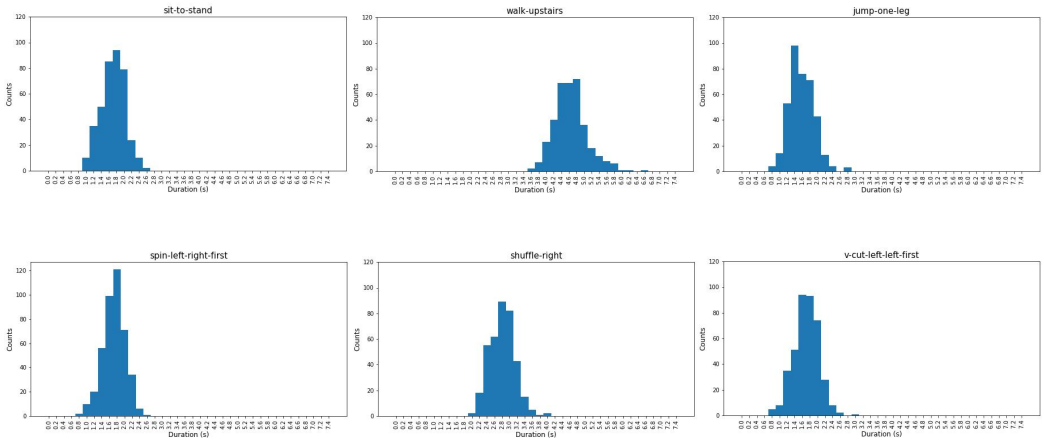


Figure 3.2 – Duration histograms of several activities in the CSL19 dataset: "sit-to-stand," "walk-upstairs," "jump-one-leg," "spin-left-right-first," "shuffle-right," and "V-cut-left-left-first."

Table 3.7 – Statistics of the segmented corpus in the CSL19 dataset. The minimum (Min.), maximum (Max.), mean, and standard deviation (std.) values are in seconds. #Seg.: number of segments.

Activity	**Min.**	**Max.**	**Mean±std.**	**#Seg.**	**Total**
sit	0.819	8.019	1.66 ± 0.58	389	00:10:47
stand	0.809	6.959	1.64 ± 0.51	405	00:11:06
sit-to-stand	1.049	2.719	1.81 ± 0.32	389	00:11:45
stand-to-sit	1.129	3.729	1.92 ± 0.35	389	00:12:28
walk	3.139	5.589	4.26 ± 0.44	400	00:28:23
walk-curve-left (90°)	2.899	6.449	4.34 ± 0.56	398	00:28:46
walk-curve-right (90°)	3.229	6.289	4.45 ± 0.50	393	00:29:10
walk-upstairs	3.789	6.729	4.76 ± 0.44	365	00:28:56
walk-downstairs	3.069	5.919	4.31 ± 0.50	364	00:26:09
spin-left-left-first	0.959	3.069	1.67 ± 0.30	380	00:10:33
spin-left-right-first	0.969	2.609	1.83 ± 0.29	420	00:12:48
spin-right-left-first	0.800	2.619	1.86 ± 0.24	401	00:12:26
spin-right-right-first	1.169	2.719	1.71 ± 0.22	400	00:11:26
run	2.319	4.119	3.15 ± 0.33	400	00:21:01
V-cut-left-left-first	0.809	3.039	1.81 ± 0.33	399	00:12:03
V-cut-left-right-first	1.019	2.709	1.88 ± 0.29	378	00:11:50
V-cut-right-left-first	0.840	2.759	1.80 ± 0.34	400	00:11:59
V-cut-right-right-first	1.209	2.649	1.84 ± 0.28	378	00:11:36
shuffle-left	1.739	3.869	2.89 ± 0.29	380	00:18:18
shuffle-right	2.089	4.159	2.91 ± 0.33	374	00:18:10
jump-one-leg	0.830	2.949	1.69 ± 0.33	379	00:10:40
jump-two-leg	0.869	3.389	1.95 ± 0.39	380	00:12:20
Total	0.800	8.019	2.54 ± 1.58	8,561	06:02:39

3.5.4 Documentation and Application

Based on the experiences and lessons learned from the CSL17 and the CSL18 recording events and the structured laboratory work, the CSL19 dataset has become a benchmark dataset that lays a solid foundation for further research work. The collected data and generated labels were not only carefully post rectified, verified, archived, and backed up, but also well documented. We can offer participants a "sensor data documentation" based on signal visualization, which provides transparency of the collected data. Appendix B exemplify four pages of such a documentation.

In a machine learning competition for college students, "4$^{\text{th}}$ Bremen Big Data Challenge" (BBDC, 2019), the CSL19 dataset served as the data for recognition. We provided 16 of the 20 sessions as data for training, and the participants used their continuously optimized recognition results of the remaining four sessions to compete for the ranking. Many teams took part in the challenge, and 61 of them submitted final results, among which 38 teams reached over 80% recognition accuracy, as demonstrated in Figure 3.3. The top five winners managed to reach over 97%, which reflects the high quality of this dataset from another perspective.

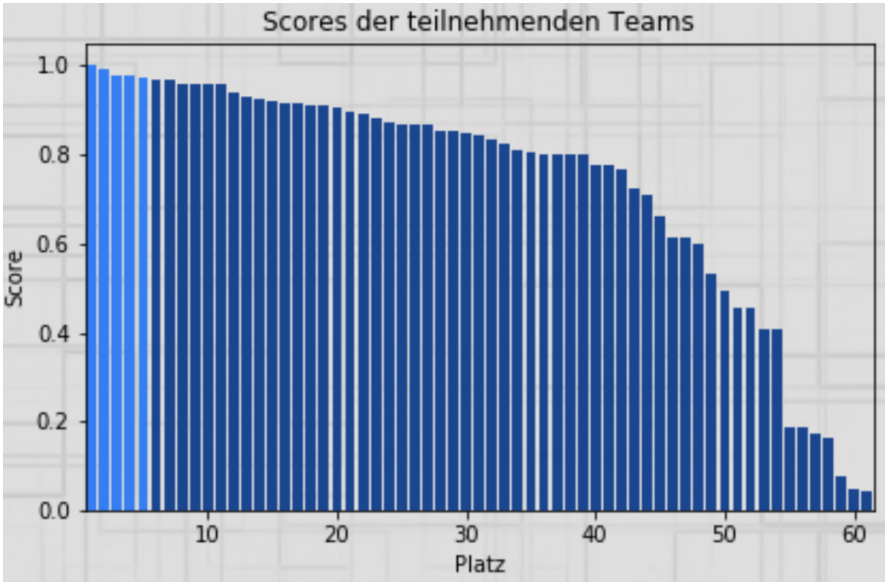


Figure 3.3 – Screenshot (inverted color) of each team's final recognition accuracy in (BBDC, 2019). "Scores" correspond to the final recognition rates.

Standing on the dataset robustness, we are working on sharing the CSL19 dataset as an open-source wearable sensor-based dataset, called CSL-SHARE (Cognitive Systems Lab Sensor-based Human Activity REcordings), hoping to contribute research materials to the researchers in the same or similar fields.

3.6 External Dataset UniMiB SHAR (UMS) and Its Activity-of-Daily-Living Subset UMS9

The *UniMiB SHAR* (UMS) dataset with 17 activities, recorded at the *University of Milano Bicocca*, was published by (Micucci et al., 2017). The dataset

of acceleration samples, designed for the HAR of *Activities of Daily Living* (ADLs) and *Fall Detection* (FD), was acquired with an *Android* smartphone. (Xue and Liu, 2021) mentions that the modeling approaches and other research aspects of FD can learn from HAR in many research scenarios. As can be seen from Table 3.8, the UMS dataset contains nine different ADLs, which are very similar to the everyday activities in our three in-house datasets (see Section 3.3—3.5), and eight types of falls, from 30 subjects of ages ranging from 18 to 60 years. Totally 9.8 hours of 11,771 samples were recorded. A triaxial smartphone accelerometer sampled at 50 Hz was used for recording, and the gravitational constant acceleration was removed post-recording. The recording with an *Android* smartphone as a sensor carrier happened in equal parts in the left and right subjects' pocket during data acquisition.

Table 3.8 – Statistics of the *UniMiB SHAR* dataset. Length: each segment's length; #Seg.: number of segments.

Activity	**Length**	**#Seg.**	**Total**
Walking	3.0 seconds	1,738	01:27:29
GoingDownStairs	3.0 seconds	1,324	01:06:38
GoingUpStairs	3.0 seconds	921	00:46:21
Running	3.0 seconds	1,985	01:39:55
Jumping	3.0 seconds	746	00:37:33
StandingUpFromSitting	3.0 seconds	153	00:07:42
SittingDown	3.0 seconds	200	00:10:04
StandingUpFromLaying	3.0 seconds	216	00:10:52
LyingDownFromStanding	3.0 seconds	296	00:14:54
All 9 ADLs		7,579	06:18:57
HittingObstacle	3.0 seconds	661	00:33:16
FallingBack	3.0 seconds	526	00:26:29
FallingBackSC	3.0 seconds	434	00:21:51
FallingForward	3.0 seconds	529	00:26:38
FallingLeft	3.0 seconds	534	00:26:53
FallingRight	3.0 seconds	511	00:25:43
FallingWithProtectionStrategies	3.0 seconds	484	00:24:22
Syncope	3.0 seconds	513	00:25:49
All 17 activities		11,771	09:48:33

To simplify the follow-up annotation work, the data collectors asked each participant to clap his/her hands before and after he/she performed the activity/fall to be recorded. Each participant was asked to wear gym trousers with front pockets to reduce background noise and place the sensors easily.

Furthermore, three recording protocols for acquiring nine ADLs were applied: 1. walking + running; 2. upstairs + downstairs + jump; 3. ascending (from sitting / from laying) + descending (sitting down / lying down). Each participant performed each protocol twice: first with the smartphone in the right pocket and then in the left.

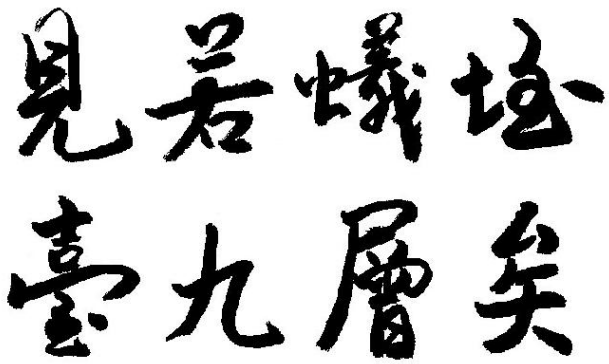
Falls have been recorded individually without protocol of sequence, following the "start clap and end clap" pattern. The participant started from a straight-up standing position. When the participant ended in a prone position, an external subject clapped for him/her to avoid unexpected movements. Each fall was repeated six times, three with the smartphone in the right trouser pocket, three in the left.

The data in the UMS dataset are automatically and uniquely segmented into three-second windows around a magnitude peak during the activities. This automatic segmentation mechanism is effortless to execute and has no demand on equipment or machine learning algorithms; however, the resulted segments are not always correct, and gait-based activities like "walk" turn out to be challenging for sequential modeling and offline training because of the indistinct number of the gait cycles during the three seconds, and the uncertainty of the initial state. The use of fixed-value window length has a good simulation for real-time HAR systems.

The full UMS dataset is applied to evaluate the feature space study (see Section 4.3), and its nine-ADL-subset UMS9 is used to verify the *Motion Units* design (see Sections 5.6 and 5.7.3).

Chapter 4

Feature Dimensionality Study and Sensor Selection



"It looks as small-scaled as an anthill,

but is actually a multi-storied terrace."

Yang Jingzhi (lived around 820), *Rhapsody of Mount Hua.*

Our HAR modeling research takes feature dimensionality reduction, feature vector stacking, and sensor selection as the premise. Besides, feature selection also helps reduce dimensionality. However, the feature selection experiments will be described after activity partitioning modeling research (see Section 5.4) due to the following reason: Numerous features take a long time to be compared, especially in the greedy forward selection experiments. Thus, we will first prepare a well-performed benchmark modeling architecture, namely phase and state partitioning (see Section 5.3), so as to run the feature selection experiments on a reasonable baseline recognizer.

Since feature selection has not been studied yet, we use the two most widely applied features, i.e., *average* and RMS (see Section 2.5.2), to study feature dimensionality and sensor selection in this chapter, as they have been proven effective in many prior research works, such as (Rebelo et al., 2013) and (Liu and Schultz, 2018).

It is noteworthy that the various preliminary studies described in this chapter do not necessarily provide the optimal solutions of the succeeded research but a strong baseline as a benchmark for the iterative process of feature dimensionality reduction, sensor selection, feature selection, and modeling research. The point of departure for this iterative process is created by the initial study depicted in the following sections.

4.1 Study on the CSL18 Dataset

The CSL18 dataset (see Section 3.4) is our first in-house collected multi-subject dataset using the complete set of selected sensors, which is different from the one-subject CSL17 dataset with a limited number of sensors. Therefore, we first studied feature dimensionality reduction, feature vector stacking, and sensor selection on the CSL18 dataset.

4.1.1 Feature Space Reduction and Primitive Experiments

The CSL18 dataset has 21 channels. If we extract n features from each channel in a window to form a feature vector, it will have $21 \times n$ dimensionality, which does not consider the stacking composition. However, not all of these features are discriminative, and some may be highly correlated to each other, which will lead to redundancy, a famous *Curse of Dimensionality* problem (Bellman et al., 1957). Given this, minimizing the redundancy between different biosignals and condensing it into a reasonably scaled feature space can not only save the time

of various subsequent computational tasks, but also improve the classification task since the training data will better cover a smaller feature space. We use *Linear Discrimination Analysis* (LDA) for feature space reduction in the following experiments, which is also applied in similar research work, such as (Uddin et al., 2008).

Reducing the feature space dimension has several benefits for HAR on the CSL18 dataset (Hartmann et al., 2020):

- GMMs can be fitted more effectively;
- More sensor-specific features can be added practically, since LDA is used to transform them into a smaller feature space.

We align the samples to states by applying the *Viterbi* algorithm, where each activities' state is regarded as the target for the aligned feature vector. Applying LDA to the baseline model on the CSL18 dataset, the person-dependent recognition accuracy, based on the model and hyperparameter values listed in Table 4.1, increases from 91.0% to the performance peak of 94.9%, when a 13-dimensional feature space is used, as shown in Figure 4.1.

Table 4.1 – Hyperparameter values for the primitive feature study on the CSL18 dataset. #: number of; dim.: dimensions.

Parameter	**Value**
Window length	10 milliseconds
Overlap length	2 milliseconds
#HMM states per activity	6
#Gaussians per state	7
Train-iteration	10
Features applied for each channel	average; RMS
Normalization	enabled
Baseline feature vector dimension	21 channels $\times$ 2 features = 42 dim.
#Cross-validation folds	10
Training data amount in each fold	90% of the whole dataset

When reducing to too few dimensions, the decrease in performance is also observable in Figure 4.1, as too little information remains. Another explanation of the steady decline of performance after 13 dimensions is too little data to cover the large feature spaces.

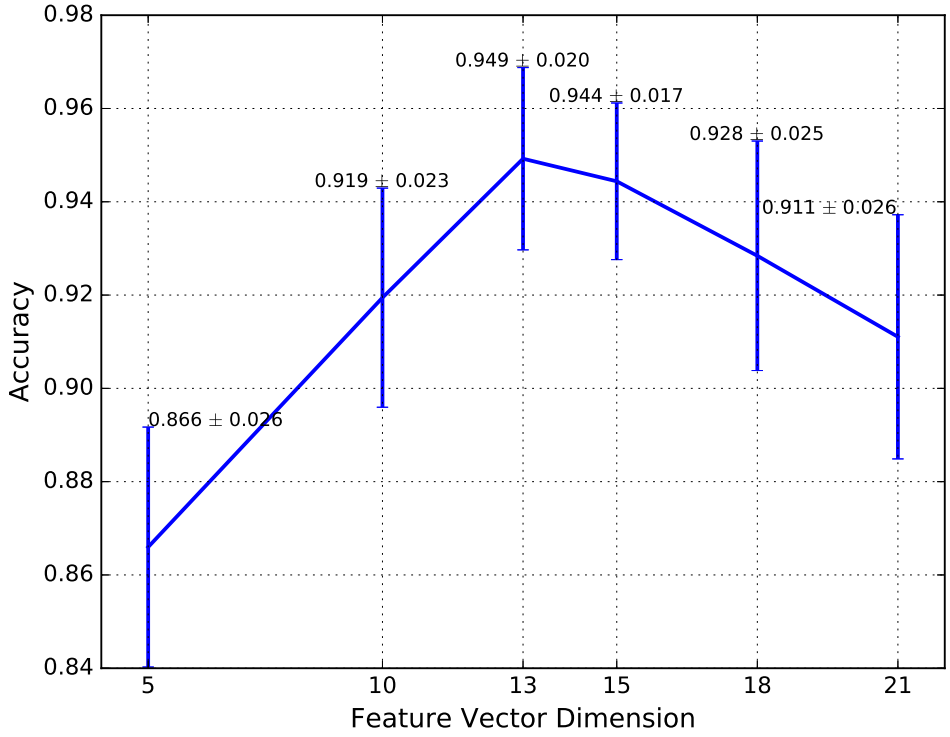


Figure 4.1 – Results of the feature space reduction experiments on the CSL18 dataset based on the original 42-dimension of features (Hartmann et al., 2020).

4.1.2 Feature Vector Stacking and Primitive Experiments

Another feasible approach to improve performance, called feature vector stacking, is to add context to each feature vector by prepending the n previous and appending the n following feature vectors, thereby increasing the time context and the vector dimension by $2n+1$ the original dimension (Hartmann et al., 2021).

Evaluated naively on the CSL18 dataset based on the same model and hyperparameter values in Table 4.1, the performance decreases with increasing context (see Table 4.2). This behavior is expected because the data sample scale is too small to train large dimensional feature vectors. That is to say, in this case, stacking feature vectors to expand context did not increase performance, but instead decreased it with respect to not stacking at all.

Table 4.2 – Results of the feature vector stacking experiments on the CSL18 dataset based on the original 42-dimension of features (Hartmann et al., 2020). dim.: dimensions.

Context	Feature vector dimension	Recognition accuracy
$n = 0$	21 channels $\times$ 2 features = 42 dim.	0.92
$n = 1$	42 dim. $\times$ $(2 \times 1 + 1)$ = 126 dim.	0.80
$n = 2$	42 dim. $\times$ $(2 \times 2 + 1)$ = 210 dim.	0.74

By applying a fixed feature space dimension of 13 (the best feature dimensionality from the feature space reduction experiments shown in Figure 4.1), the performance of the feature vector stacking experiments on the CSL18 dataset was not improved (see Figure 4.2). The results between different context sizes are not significant, as a statistical analysis via T-Test indicates.

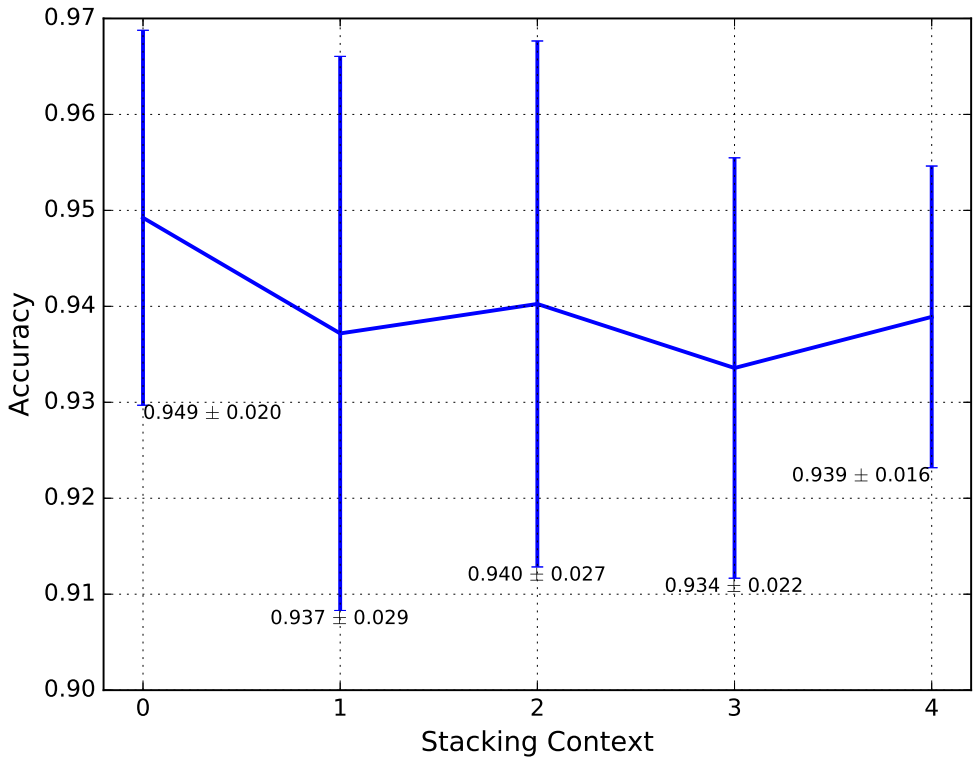


Figure 4.2 – Results of the feature vector stacking experiments on the CSL18 dataset obtained on a local optimum with a 13-dimensional feature space (Hartmann et al., 2020).

4.1.3 Sensor Selection

As introduced in Chapter 3, based on the conventional combination of sensors for HAR applied in the CSL17 dataset, i.e., triaxial accelerometers, EMG sensors, and the electrogoniometer, the CSL18 dataset brought four additional types of sensors into the acquisition: gyroscopes, the airborne microphone, the piezoelectric microphone, and the force sensor. Nevertheless, some of these newly introduced sensors were finally removed in the succeeding research from the dataset based on a series of investigations and experiments.

The force sensor (see Figure 2.4⑥) channel became the first candidate to be removed due to the following reasons:

- The physical limitation of the sensor placement: The sensor's positioning is very person-specific, and the shifting happens in different ways for different subjects. Furthermore, due to the positioning instability, the force sensor was easily broken;
- The unstable measurement results: Even if the force sensor was correctly placed in some recording sessions and no damage occurred, the data analysis of the acquired signals still reflected poor stability and consistency.

It must be clarified that the above two points do not mean the force sensor is of low quality. Actually, it performed very well in pre-tests. Moreover, there is no evidence that the force sensor is not suitable for HAR systems. The problems listed above are mainly because the scheme of the force sensor location and the bandage integration do not perfectly cooperate with the selected sensor product. In the future, if we have a better solution between the sensor product and the integration scheme, the value of force sensors for HAR can still be further researched.

The remaining six types of sensors were compared through backward selection experiments with cross-validation. Based on the feature dimensionality study framework introduced in Sections 4.1.1 and 4.1.2, a series of experiments with different parameter setups were executed and provided similar results of the sensor performance. Table 4.3 lists the model and hyperparameter values of one representative experiment, and Figure 4.3 shows the correspondent experimental results.

Some clues can be observed from Figure 4.3 on the CSL18 dataset:

Table 4.3 – Hyperparameter values of one representative sensor selection experiment on the CSL18 dataset. #: number of.

Parameter	Value
Window length	10 milliseconds
Overlap length	0
#HMM states per activity	4
#Gaussians per state	8
Context of stacking	$n = 2$ ($2 \times 2 + 1 = 5$ frames)
#Cross-validation folds	10
Training data amount in each fold	90% of the whole dataset

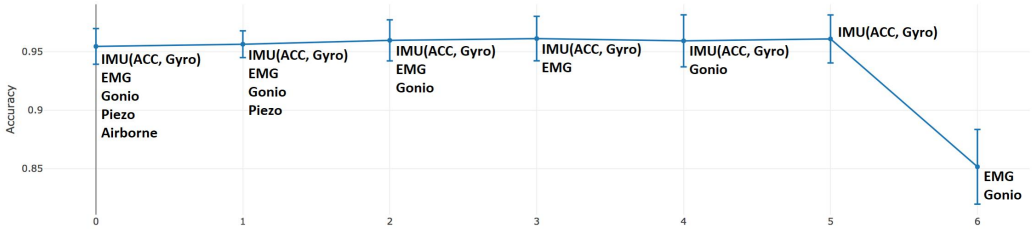


Figure 4.3 – Recognition results of the representative sensor selection experiment on the CSL18 dataset. Each vertical bar is the mean recognition accuracy of ten cross-validation repetitions.

- The complete set, i.e., the combination of all sensors (except force sensor), does not provide the best performance;
- IMU sensors (accelerometers and gyroscopes) make the main contribution to the recognition;
- EMG sensors and the electrogoniometer do not work as well as IMUs;
- Acoustic sensors (the piezoelectric microphone and the airborne microphone) do not significantly contribute to differentiating between activities.

Although the primitive sensor selection results do not favor the EMG sensors and the goniometer, based on the overall research and multiple literature, EMG and electrogoniometer signals are not redundant in many cases. For example, our real-time recognizer ASKED (see Section 5.1.3) confirmed their effectiveness for distinguishing typical activities such as "stand" versus "sit." So, keeping these two types of sensors in our further research is necessary, which also indicates that the study of sensor performance here is only meaningful to the activities involved in this dissertation (i.e., the activities included

in the CSL17, the CSL18, the CSL19, and the UMS9 datasets). We should be open to HAR's sensor solutions: Sensors that are not commonly applied in wearable sensor-based HAR so far, such as acoustic sensors, may potentially recognize some human activities well. For example, for the recognition of the activity "clap hands," we can imagine the outstanding capability of the acoustic sensors.

As the experiments have proved that inertial sensors contribute the most to the recognition, we were still interested in the cooperation between different types and positions of the inertial sensors.

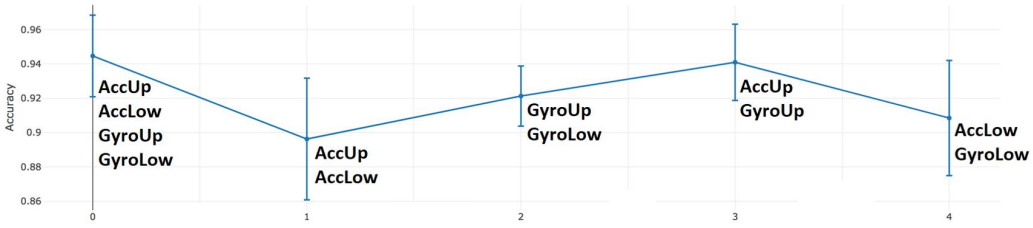


Figure 4.4 – Results of performance comparison experiments of both IMU sensors in different positions. Up: thigh; Low: shank.

From Figure 4.4, we can draw some conclusions about the two types of IMU sensors applied in the CSL18 dataset:

- The angular acceleration offer better recognition results than the linear acceleration;
- IMU Sensors placed on the thigh are more helpful for recognition than on the shank.

What needs to be pointed out is that all the above experimental results on sensor selection are based on the CSL18 dataset with fixed number of HMM states. However, the iterative process of sensor selection based on improved activity modeling and hyperparameters deserves further study in the future, not included in this dissertation. The purpose of the sensor selection and the comparison experiments on the CSL18 dataset is to provide evidence of the sensor effectiveness for the subsequent larger-scaled CSL19 data collection (see Section 3.5).

4.1.4 Joint Feature Dimensionality Study

Classic evaluation criteria like accuracy may not be perfectly applicable to our joint feature dimensionality study due to the following two reasons:

- We will use different datasets, two in-house (CSL18 and CSL19) and one external (UMS), to ensure the joint feature dimensionality study's comprehensiveness. Each dataset has a different set of activities;
- As a result of the protocol design and the uncertainty in the data collection experiments, most of the activity's number of occurrences are different from the others in each dataset.

In order to generate balanced results with higher reference values on the experiments combining the feature vector stacking and feature space reduction, we use the arithmetic mean of all activity recognition accuracies (macro-averaging) to compare the feature vector stacking and feature space reduction results.

In the joint study of feature space reduction and feature vector stacking, not only person-dependent but also person-independent study were included. In contrast to the former separate reduction and stacking study on the CSL18 dataset in Sections 4.1.1 and 4.1.2, we optimized the model and the hyperparameters based on the experimental analysis including cross-validation with successive grid search, as shown in Table 4.4. The parameters are optimized with a person-independent evaluation rather than a dependent one, which results in a more robust recognizer and advances the experimental results to be comparable with the other two datasets' results (see Sections 4.2.2 and 4.3).

Table 4.4 – Hyperparameter values for the joint feature dimensionality study on the CSL18 dataset. #: number of.

Parameter	**Value**
Window length	30 milliseconds
Overlap length	6 milliseconds
#HMM states per gait-cycle	5
#HMM states for sit/sit-to-stand/stand/stand-to-sit	1
Tuning of number of Gaussians per state	*Split-and-Merge*
Train-iteration	10
Features applied for each channel	average; RMS
Normalization	enabled
Baseline feature vector dimension	40 dimensions
#Cross-validation folds, person-independent	4
#Cross-validation folds, person-dependent	5
#Training data amount in each fold, person-dependent	80%

The HMM topology uses five states for each gait cycle in an activity. "Run," for example, contains three full cycles and therefore uses 15 states. "Sit," "stand," and the transitions between them are modeled using a single state. The optimized baseline achieves $63.8 \pm 8.5\%$ macro-average accuracy in an independent evaluation, 30-millisecond windows with 6-millisecond overlap, and normalization enabled. The force sensor channel was removed from the sensor group as the signal quality was not stable (see Section 4.3), making the baseline feature dimension equal to 40 instead of 42.

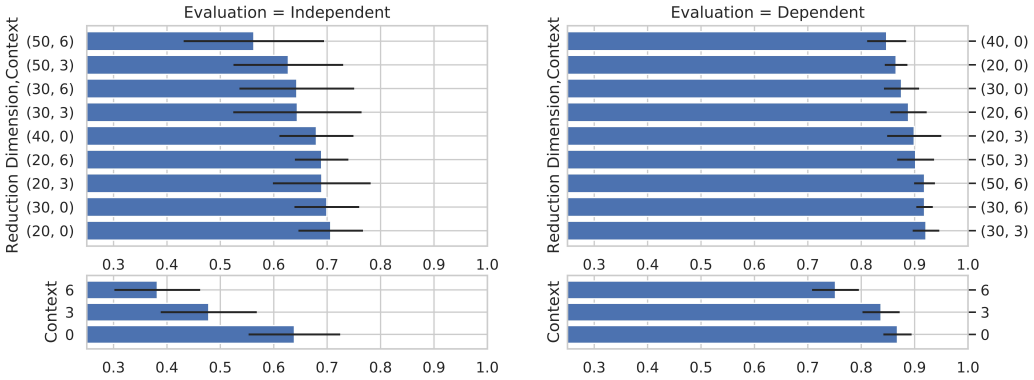


Figure 4.5 – Results (recognition accuracies) for joint feature vector stacking and feature space reduction experiments on the CSL18 dataset. Left column: person-independent evaluation; right column: person-dependent evaluation; upper row: the combination of stacking and reduction (dimension target, stacking context); lower row: stacking alone. Without stacking, there is no reduction to 50 dimensions, but a transformation of the original 40 dimensions (Hartmann et al., 2021).

As shown in Figure 4.5, in the person-independent evaluation, the best performances are achieved without stacking at $70.6 \pm 6\%$ macro-average accuracy when reducing to a 20-dimensional feature space. Based on a statistical analysis via T-Test, this result is not significantly different from other parameter combinations.

Compared to the experimental results of the CSL19 dataset (see Figure 4.7), the recognition accuracies of person-independent experiments on the CSL18 dataset are significantly worse than the person-dependent. The reason is that we did not specify the number of steps during the CSL18 data acquisition, mainly for the "walk," "run," "walk-curve," and "V-cut" activities, due to a compromise in the collaborative project research, as introduced in Section 3.4.1.

A closer look at the performance increases per activity shown in Figure 4.6 reveals that the walking and running derivatives (e.g., "walk-curve-left (90°)," "walk-curve-right (90°)," "walk-upstairs," "walk-downstairs," "run," "V-cut-left," and "V-cut-right") are recognized better, while the static activity "sit" is much worse recognized.

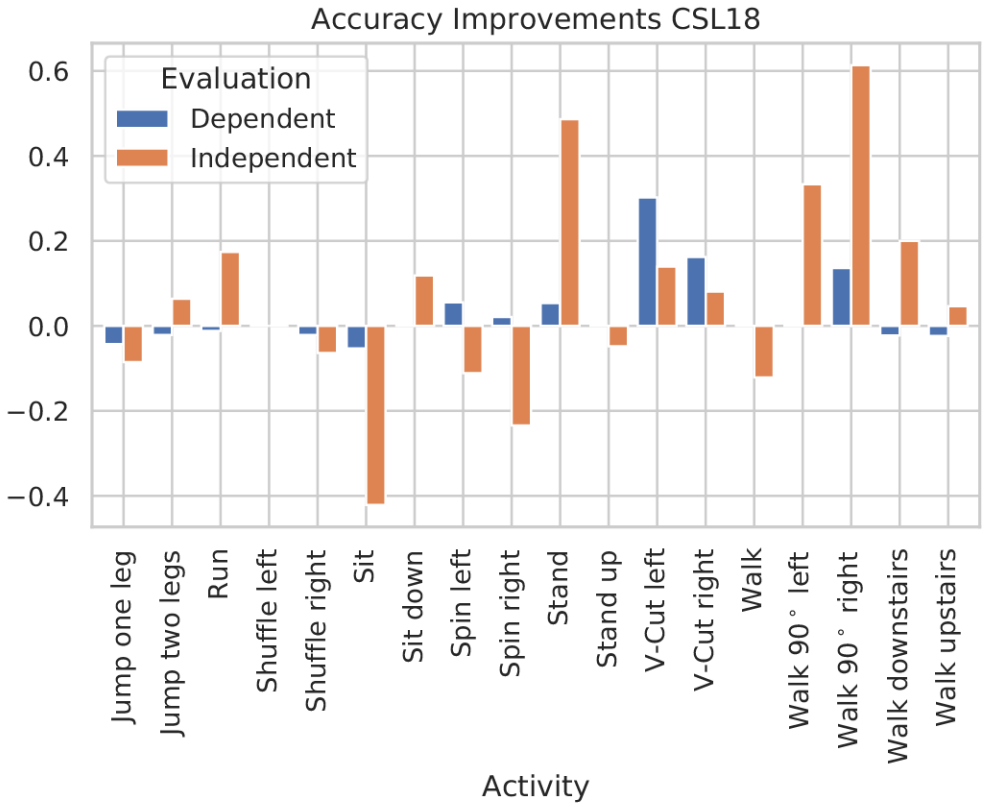


Figure 4.6 – Relative recognition accuracy improvements in percentage points between the baseline and the reduction-based recognizer on the CSL18 dataset (Hartmann et al., 2021).

4.2 Study on the CSL19 Dataset

As introduced in Section 3.5.1, the CSL19 dataset involves more subjects and more activities than the CSL18 dataset. In the CSL19 data acquisition, we used the same complete set of sensors as the CSL18 dataset. Because the CSL19 dataset will serve as the benchmark dataset for subsequent HAR modeling studies, including *Motion Units* (see Chapter 5), we first retrench

the number of redundant sensors and sensor channels based on the previous sensor performance study results (see Section 4.1.3) and the information obtained from the application of the dataset.

4.2.1 Sensor and Channel Selection

Based on the exact reason described in Section 4.1.3, the force sensor was firstly removed. Afterward, we removed the two acoustic channels because they proved to be unable to provide effective information in the sensor selection experiment of the CSL18 dataset, and they also did not reflect their value in various experiments of the CSL19 dataset. Besides, the piezoelectric microphone often output unstable signals due to the physical limitation of the sensor placement, similar to the force sensor problem. However, it does not mean that acoustic sensors are not suitable for an HAR system, but merely in our experiments so far there is no evidence to approve their value.

In addition to the sensor selection, there is also a channel selection problem for the electrogoniometer, as we use a biaxial one.

According to our integration solution, the electrogoniometer measures angles in two planes, the horizontal and the sagittal. Possible motions of the knee are limited to flexion/extension and rotation. Our measurements indicate no significant angle differences on the horizontal plane. Therefore, the angular measurements of the horizontal plane can be discarded. Besides, results from the "4$^{\text{th}}$ Bremen Big Data Challenge" (BBDC, 2019) (see Section 3.5.4) suggest that the inclusion of the horizontal plane information reduced recognition results, negating its applicability from a practical perspective. Therefore, in the follow-up research on the CSL19 dataset, we use only one of the two electrogoniometer channels which senses the angles on the sagittal plane.

4.2.2 Joint Feature Dimensionality Study

Table 4.5 lists all optimized hyperparameters for the CSL19 dataset based on experimental analysis.

In Section 4.2.1, we introduced the reasons why several sensors and channels were removed from the CSL19 dataset. As a result, the CSL19 dataset contains 17 channels: two triaxial accelerometers, two triaxial gyroscopes, four EMG sensors, and one electrogoniometer. The optimized HMM topology for the CSL19 dataset uses the same number of states as in the joint study of the CSL18 dataset (see Section 4.1.4). For the initial hyperparameters, a *Hamming* window with 20-millisecond overlap, enabled normalization, and 10

HMM train-iterations are chosen based on the previous work (Hartmann et al., 2020). The window size is determined in the first experiment and does not require an initial value. As can be found in Table 4.5, after optimization, the best parameters are 100-millisecond *Hamming* windows with 50-millisecond overlap, 10 HMM train-iterations, and normalization enabled. The baseline achieves a $93.7 \pm 1.4\%$ macro-average accuracy in a person-independent leave-one-person-out cross-validation experiment and $97.8 \pm 0.2\%$ macro-average accuracy in a stratified person-dependent five-fold cross-validation experiment (see Figure 4.7) (Hartmann et al., 2021).

Table 4.5 – Hyperparameter values for the joint feature dimensionality study on the CSL19 dataset. #: number of.

Parameter	**Value**
Window length	100 milliseconds
Overlap length	50 milliseconds
#HMM states per gait-cycle	5
#HMM states for sit/sit-to-stand/stand/stand-to-sit	1
Tuning of number of Gaussians per state	*Split-and-Merge*
Train-iteration	10
Features applied for each channel	average; RMS
Normalization	enabled
Baseline feature vector dimension	$17 \times 2 = 34$ dims
#Cross-validation folds, person-independent	20
#Cross-validation folds, person-dependent	5
#Training data amount in each fold, person-dependent	80%

As shown in Figure 4.7, the best performance in an independent evaluation is achieved using a stacking context of three and reducing to a 50-dimensional feature space at $93.3 \pm 3.9\%$ accuracy. Notably, this is not higher than the $93.7 \pm 1.4\%$ achieved without reduction, which is also not significant as a statistical analysis via T-Test indicates. A similar observation can be made in the person-dependent evaluation: $97.8 \pm 0.2\%$ without and $97.4 \pm 0.3\%$ with the reduction.

Figure 4.8 shows that while the reduction-based recognizer can better distinguish "walk" and "walk-curve-left/right (90°)," "stand" is confused more often, resulting in a similar overall performance.

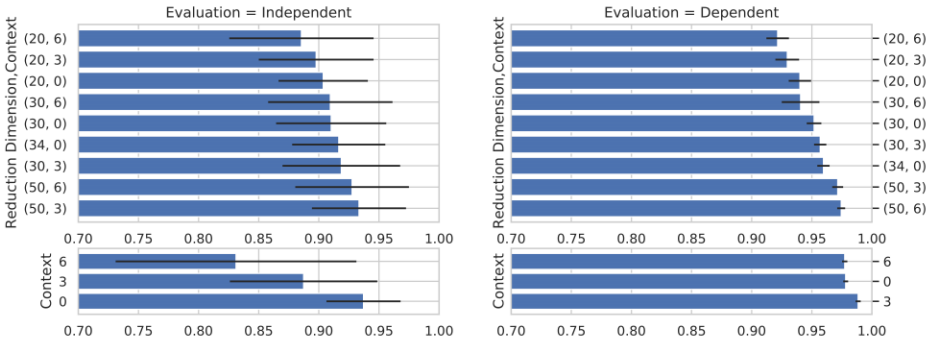


Figure 4.7 – Results (recognition accuracies) for joint feature vector stacking and feature space reduction experiments on the CSL19 dataset. Left column: person-independent evaluation; right column: person-dependent evaluation; upper row: the combination of stacking and reduction (dimension target, stacking context); lower row: stacking alone. Without stacking, there is no reduction to 50 dimensions, but a transformation of the original 34 dimensions (Hartmann et al., 2021).

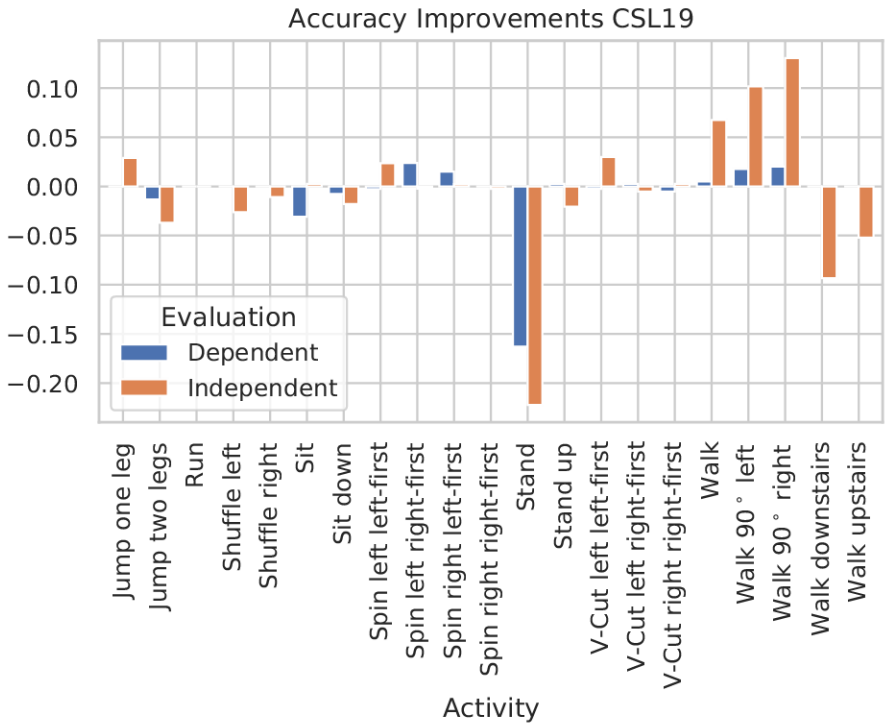


Figure 4.8 – Relative recognition accuracy improvements in percentage points between the baseline and the reduction-based recognizer on the CSL19 dataset (Hartmann et al., 2021).

4.3 Joint Feature Dimensionality Study on the UMS Dataset

Experiments were also executed on the external UMS dataset to provide more references. The activities containing gait cycles are modeled similarly to the CSL19 activities, with five states for each cycle. All "fall" activities are modeled with ten states, as this performed best in the person-independent cross-validation experiments.

The UMS dataset contains fewer samples per segmented activity due to the lower sampling rate. Therefore, a single grid search is feasible and executed. The best parameters are 400-millisecond *Hamming* windows with 320-millisecond window overlap, 30 HMM train-iterations, and normalization enabled. The baseline achieves a $59.7 \pm 8.6\%$ macro-average accuracy in a person-independent leave-one-person-out cross-validation experiment (Hartmann et al., 2021).

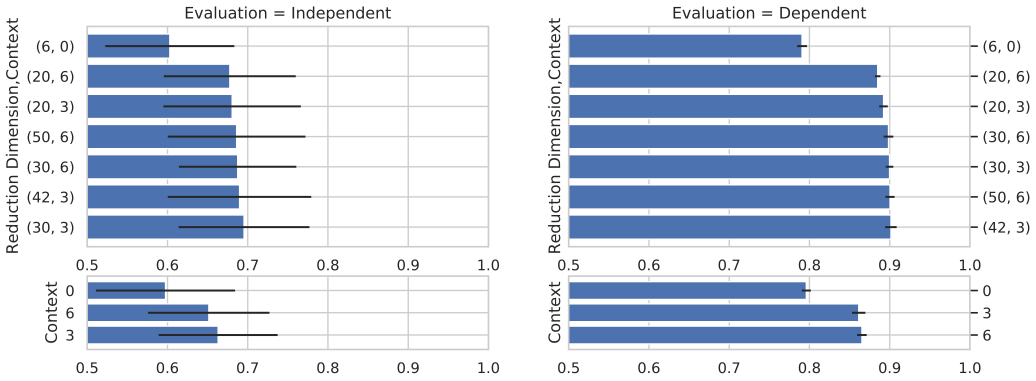


Figure 4.9 – Results (recognition accuracies) for joint feature vector stacking and feature space reduction experiments on the UMS dataset. Left column: person-independent evaluation; right column: person-dependent evaluation; upper row: the combination of stacking and reduction (dimension target, stacking context); lower row: stacking alone. Without stacking, there is no reduction to 20, 40, or 50 dimensions, but a transformation of the original six dimensions (Hartmann et al., 2021).

The feature space reduction is evaluated on a grid ranging from 0 to 6 frame-stacking context and a target dimension from 20 to 50. Figure 4.9 brings out the experimental results. Notably, stacking alone already increases performance, and the reduction extends this. Similar to the CSL19 dataset, the target dimension influences performance more than the stacking context.

The best performing parameter combination with a target dimension of 30 and a context of 3 performs at $69.5 \pm 8.1\%$ macro-average accuracy in a leave-one-person-out cross-validation experiment. Compared to the $59.7 \pm 8.6\%$ accuracy in the baseline, this is an improvement by 10 percentage points. This difference is also apparent in the performance improvements for almost every activity, as depicted in Figure 4.10.

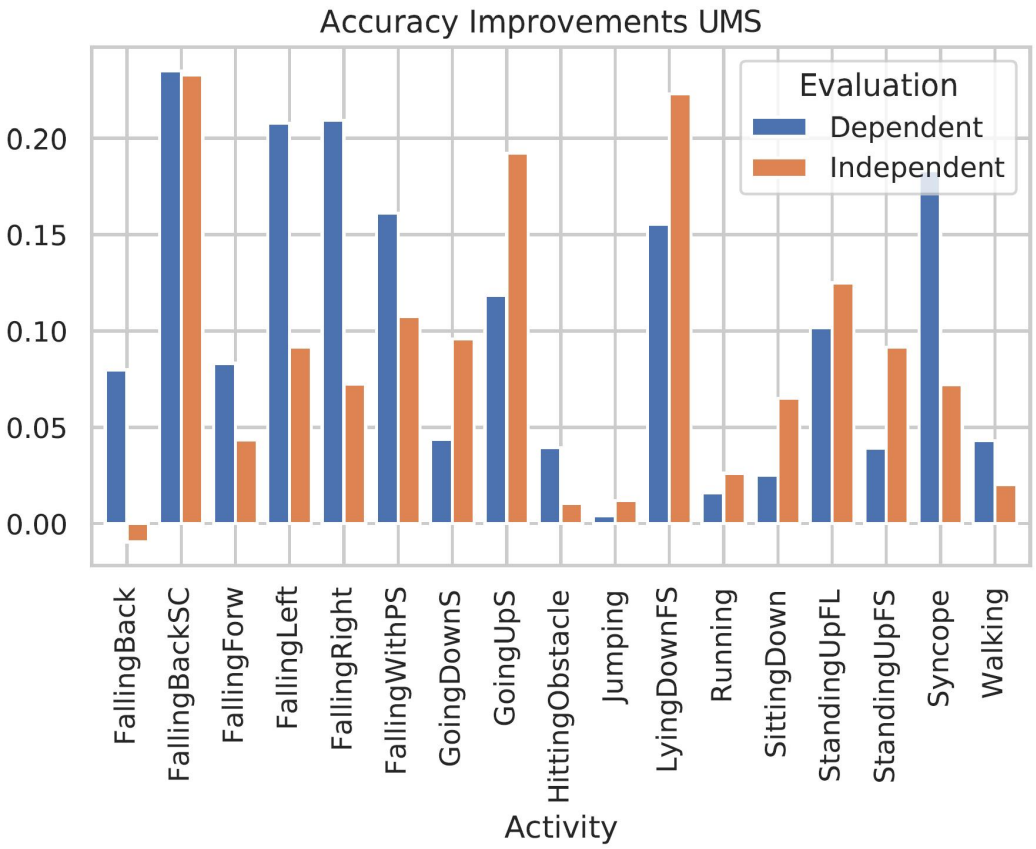


Figure 4.10 – Relative recognition accuracy improvements in percentage points between the baseline and the reduction-based recognizer on the UMS dataset (Hartmann et al., 2021).

As can be found obviously in Figure 4.9, it is also worth mentioning that compared to the CSL19 dataset, the recognition results of person-independent experiments of the UMS dataset are much worse than the person-dependent, no matter for the baseline or the stacking/reduction. Possible reasons are as follows:

- In addition to eight ADLs similar to those in the CSL19 dataset, UMS also contains nine types of falls. Generally, the falling activities increase the difficulty of person-independent recognition;
- All falls were primitively modeled by ten HMM states in the feature dimensionality study, a temporary topology based on the highest recognition accuracy of the "fixed number of states" modeling (see Section 5.2.1). We are going to study the fall modeling improvement, but it is not the focus of this dissertation;
- Compared with the CSL19 dataset adopting protocol-for-pushbutton semi-automatic segmentation, which makes as much as possible to retain all subjects' data for the same activity entirely, UMS adopts a fixed three-second segmentation process for all activities, which potentially loses some helpful information and magnifies the signal difference between subjects for the same activity. Significantly, the activities containing complex body movements with considerably longer duration, such as "LyingDownFromStanding" and "StandingUpFromLaying," are not ideally recognized in the person-independent experiments compared to the other activities, for which we can find some clues from UMS9 confusion matrices and criteria tables of the recognition results in Section 5.6 (Figures 5.26, 5.28, 5.30, and 5.32, as well as Tables 5.10, 5.12, 5.14, and 5.16).

4.4 Conclusion

Feature space reduction and feature vector stacking are not only an important research topic of HAR, but also an additional inspection process of data quality and model parameters. We applied three datasets, CSL18, CS19, and UMS, and their respective recognizers to study the feature dimensionality, between which there are several differences, such as:

- The number of participants: CSL18: 4 participants; CSL19: 20 participants; UMS: 30 participants;
- The types of activities: CSL18 and CSL19: ADLs + sport; UMS: ADLs + falls;
- The activity modeling topologies;
- The evaluated window length: CSL18: 30 milliseconds; CSL19: 100 milliseconds; UMS: 400 milliseconds;

- The window overlaps: CSL: 6 milliseconds; CSL19: 50 milliseconds; UMS: 320 milliseconds.

Across all three datasets, the feature space reduction with an LDA did improve recognition accuracy for several activities. The resulting overall performance is either similar or significantly higher compared to the baseline. Table 4.6 gives a summary of the evaluation results.

Table 4.6 – Summary of the feature dimensionality experimental results: the baselines and the reduction-based recognizers for both person-independent and person-dependent evaluation on the CSL18, the CSL19 and the UMS datasets (Hartmann et al., 2021). Data.: Dataset.

Model / **Data.**	Person-independent		Person-dependent	
	Baseline	Reduction	Baseline	Reduction
CSL18	$63.8 \pm 8.5\%$	$70.6 \pm 6.0\%$	$86.7 \pm 2.6\%$	$92.1 \pm 2.4\%$
CSL19	$93.7 \pm 1.4\%$	$93.3 \pm 3.9\%$	$97.8 \pm 0.2\%$	$97.4 \pm 0.3\%$
UMS	$59.7 \pm 8.6\%$	$69.5 \pm 8.1\%$	$79.6 \pm 0.5\%$	$90.1 \pm 0.7\%$

The best performance on the CSL19 dataset with $93.7 \pm 1.4\%$ person-independent accuracy fits in with the over 90% accuracies reported in other works (Rebelo et al., 2013), (Demrozi et al., 2020), (Lara and Labrador, 2012) even though distinguishing more classes. The results on the UMS dataset are directly comparable to current research and better or on par. For example, the person-independent macro-average accuracy of $69.5 \pm 8.1\%$ is significantly higher than the 56.53% in (Micucci et al., 2017), and the absolute person-independent accuracy is on par with the 77.03% accuracy previously reported (Li et al., 2018).

As the results reveal, LDA transformation does contribute to better overall recognition performance, most notably in the UniMiB and the CSL18 datasets. The reduction was beneficial to recognizing most activities, with the notable exception of the two static activities sit and stand in the CSL19 dataset, which is likely impacted by the normalization.

These experiments show that the feature space reduction using an LDA trained with HMM forced alignment targets can significantly improve the individual recognition accuracy of most activities and overall accuracy on the CSL18 and UMS datasets, while retain on the CSL19 datasets. On the CSL19 dataset, the overall accuracy decreases very slightly, which may be due to a strong benchmark. The purpose of feature dimensionality research in this chapter is to provide a strong baseline for the subsequent HAR research. We will continue to use most of the hyperparameters optimized based on this chapter's

feature dimensionality study in the subsequent chapter's activity modeling experiments. In addition, based on the recognition accuracies listed in Table 4.6, we will not apply feature vector stacking and feature space reduction on the CSL19 modeling experiments, while for the UMS experiments, we will do the opposite.

Chapter 5

Human Activity Modeling and Experiments: Towards Model Generalization of Motion Units

"*The extension of knowledge lies in*

exploring the core and ingredients of things."

Book of Rites, Chapter 42: Great Learning (Warring States period).

As introduced in Section 2.6, many time series modeling technologies have proven their capabilities in the research domain of human activity modeling, such as DNN and HMM. Both modeling technologies focus on the typical HAR problem, in which inputs are multichannel time series biosignals recorded from a set of sensors and outputs are predefined human activities.

The research on DNN usually aims to refine the automatically learned features as the higher-level abstract representation of low-level time series signals through the deep architecture (Wang et al., 2019), (Yang et al., 2015). In neural networks, the training and recognition procedure of a target activity must be divided into layers, which are often uninterpretable. In many cases, researchers do not know each layer's specific physical meaning. In contrast, the concept of "states" in the HMM definition-tuple (see Section 2.6.2) can present a better explanatory power of the activities' internal structure, which is the research focus of human activity modeling in this chapter.

5.1 Single-State HMM Modeling and Experiments

First, we consider and practice the most straightforward and most convenient HMM modeling method: single-state HMM.

5.1.1 Topology

The single-state HMM topology models each activity with a single HMM state, i.e., the fundamental units of model training and recognition are "activities" themselves. Figure 5.1 illustrates three single-state HMM modeling examples of the activities "walk," "sit-to-stand," and "stand-to-sit."

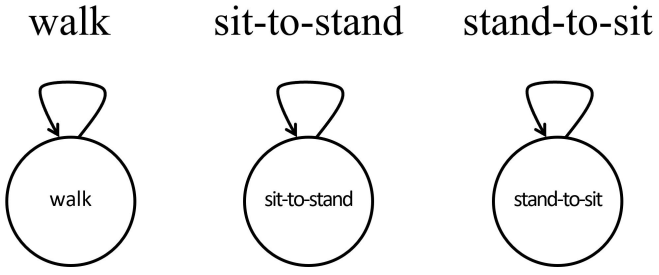


Figure 5.1 – Single-state HMMs for the activities "walk," "sit-to-stand," and "stand-to-sit."

5.1.2 Experiments on the CSL17 Dataset

The experiments on the CSL17 dataset mainly focus on the following purposes:

- To verify the applicability of the HAR research pipeline introduced in Section 2.1.3 by following the streamline of equipment selection, data collection, segmentation, annotation, feature extraction, model training, and recognition;
- To examine the robustness of the ASK software introduced in Section 2.3.1 by collecting multichannel biosignals continuously, as well as the completeness and correctness of the recorded experimental data, the segments, and the annotated labels;
- To test a baseline HMM-based HAR research environment in the ASK software, including data preparation, feature extraction, model training, and decoding, based on the in-house developed *BioKIT* decoder introduced in Section 2.6.4, as the preparation of the subsequent modeling research;
- To practice and verify the single-state HMM modeling of human activities via model training and recognition experiments.

In the pilot experiments of the CSL17 dataset, each activity is modeled by one HMM state, where nine Gaussians model each state. This setup gave the best results on a tuning test set. To evaluate the recognition accuracy, we performed a ten-fold cross-validation, i.e., we applied ten folds with, each time, 90% of data for training the GMM and 10% for testing the resulting models. For proof-of-concept, the CSL17 data acquisition was limited to one object, which determines that only person-dependent experiments can be carried out.

A mean normalization is applied to the acceleration and EMG signals to reduce the impact of Earth acceleration and to set the EMG signals' baseline to zero. The EMG signal is rectified, a widely adopted signal processing method. Before feature extraction, the signals are windowed using a rectangular window function with overlapping windows. Based on the best results of initial experiments, we chose a window length of 400 ms with a window overlap of 200 ms. We applied two features to each frame: *average* for accelerometer/electrogoniometer channels and RMS for EMG channels.

Figure 5.2 shows the confusion matrix of the recognition results on the CSL17 dataset using its complete sensor set. Table 5.1 gives the criteria precision, recall, F-score, and global accuracy of the recognition results.

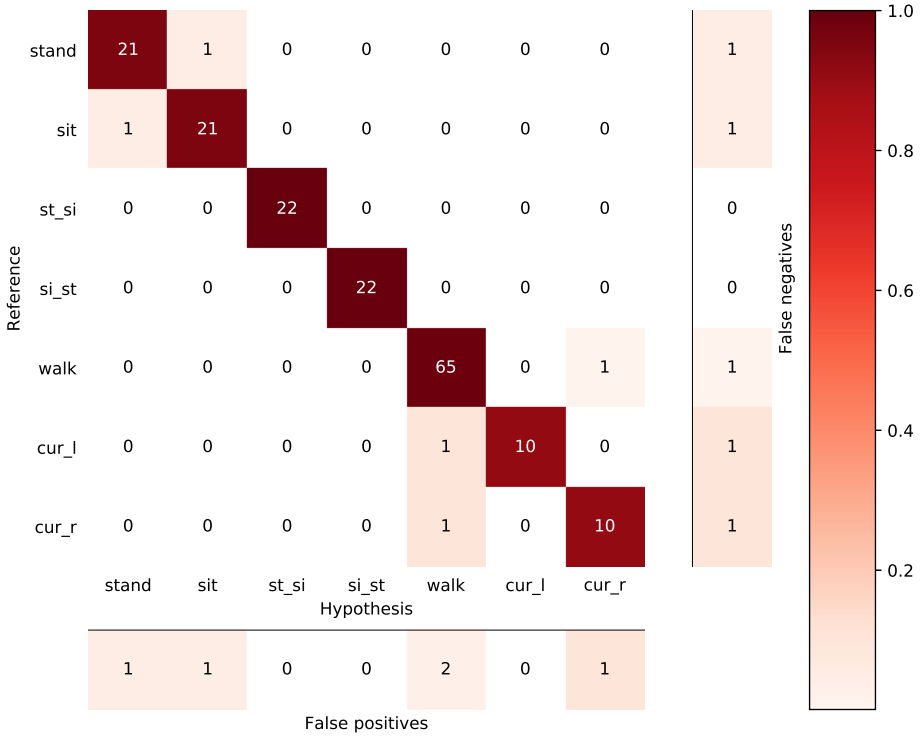


Figure 5.2 – Confusion matrix for the recognition results of the complete sensor set on the CSL17 dataset in a number of recognized activities. The color illustrates the percentage of recognition accuracy, while the numbers represent the absolute number of recognized activities. st_si: stand-to-sit; si_st: sit-to-stand; cur_l: walk-curve-left; cur_r: walk-curve-right (Liu and Schultz, 2018).

Table 5.1 – Criteria of the recognition results using the complete sensor set on the CSL17 dataset (Liu and Schultz, 2018).

Activity	Precision	Recall	F-Score
stand	0.95	0.95	0.95
sit	0.95	0.95	0.95
stand-to-sit	1.00	1.00	1.00
sit-to-stand	1.00	1.00	1.00
walk	0.97	0.98	0.98
walk-curve-left	1.00	0.91	0.95
walk-curve-right	0.91	0.91	0.91
Global accuracy: 0.97			

The activity recognition results for the small-scale CSL17 dataset have been an initial encouraging step. The global recognition accuracy reached 97%. The activities "sit-to-stand" and "stand-to-sit" are correctly recognized. "Stand" and "sit" give mixed results, for they are stable body gestures. The activities "walk," "walk-curve-left," and "walk-curve-right" exhibit to be confusable, which corresponds to our expectation.

We also performed experiments on single sensor setups. Figures 5.3—5.5 illustrate the confusion matrices for the recognition results, and Table 5.2 summarizes the single-sensor-based global accuracy.

Table 5.2 – Sensor-based recognition accuracy on the CSL17 dataset (Liu and Schultz, 2018).

Sensor	Global accuracy
Accelerometer	0.93
EMG	0.63
Electrogoniometer	0.74
All	0.97

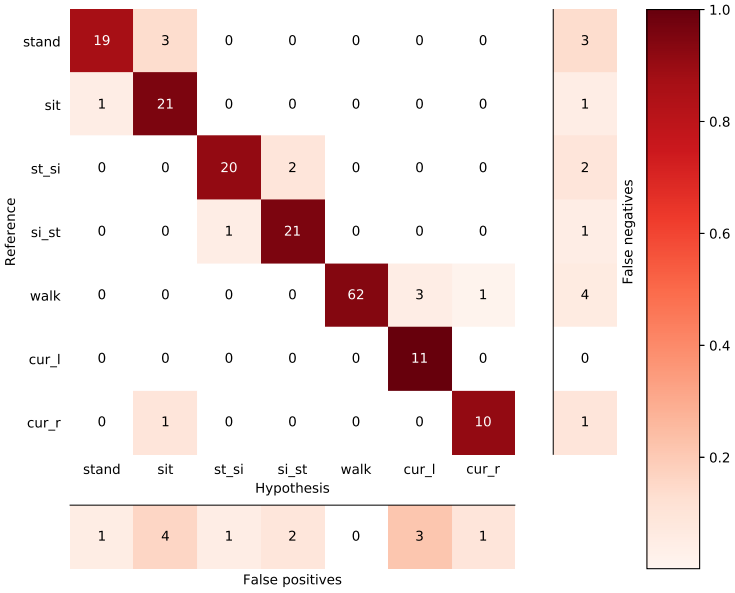


Figure 5.3 – Confusion matrix for recognition results of two accelerometers on the CSL17 dataset in a number of recognized activities. The color illustrates the percentage of recognition accuracy, while the numbers represent the absolute number of recognized activities. st_si: stand-to-sit; si_st: sit-to-stand; cur_l: walk-curve-left; cur_r: walk-curve-right (Liu and Schultz, 2018).

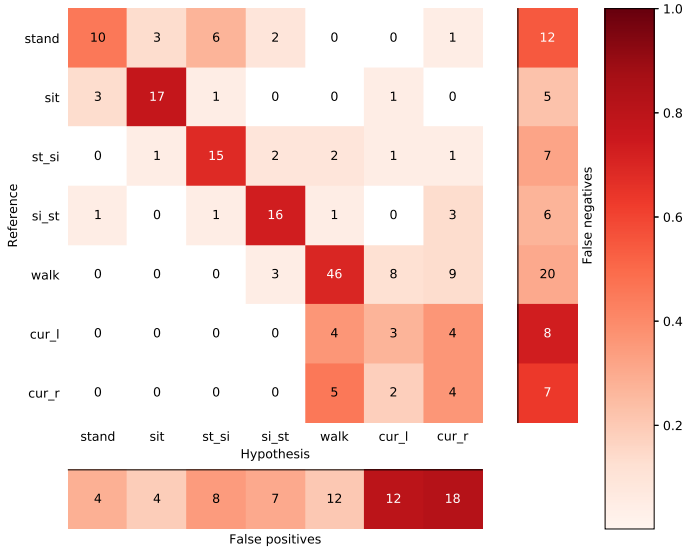


Figure 5.4 – Confusion matrix for recognition results of four EMG sensors on the CSL17 dataset in a number of recognized activities. The color illustrates the percentage of recognition accuracy, while the numbers represent the absolute number of recognized activities. st_si: stand-to-sit; si_st: sit-to-stand; cur_l: walk-curve-left; cur_r: walk-curve-right (Liu and Schultz, 2018).

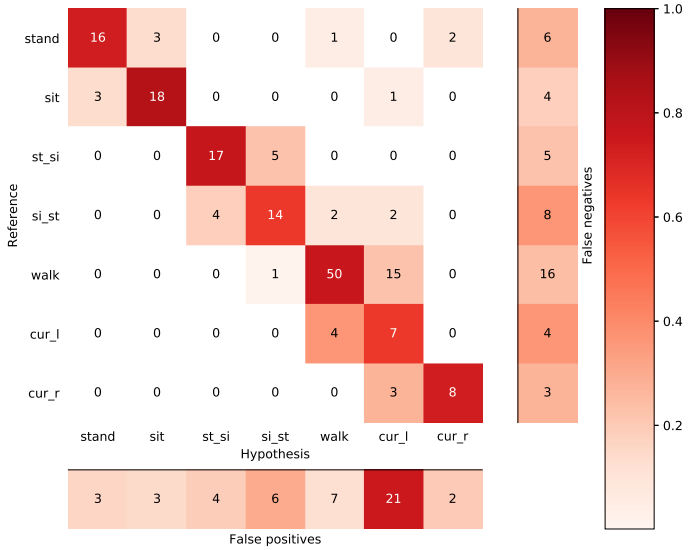


Figure 5.5 – Confusion matrix for recognition results of one electrogoniometer on the CSL17 dataset in a number of recognized activities. The color illustrates the percentage of recognition accuracy, while the numbers represent the absolute number of recognized activities. st_si: stand-to-sit; si_st: sit-to-stand; cur_l: walk-curve-left; cur_r: walk-curve-right (Liu and Schultz, 2018).

Results from Table 5.2 indicate that the use of accelerometers alone achieves an accuracy of 93%, outperforming the single-sensor results when using EMG sensors or the electrogoniometer alone. However, the latter two still enhance performance if we compare all sensors' results and accelerometers alone.

The experiments of single-HMM modeling on the CSL17 dataset have an irreplaceable role and a significance for subsequent research, given that:

- The practicability of our HAR research pipeline has been confirmed;
- The process of collecting data with the ASK software is stable and straightforward;
- The segmentation and annotation with the protocol-for-pushbutton mechanism are efficient, and the resulted timestamps and labels are directly usable;
- The baseline HAR research environment developed based on the *BioKIT* decoder runs smoothly;
- The quality of the recorded CSL17 dataset has been guaranteed under the verification of experimental results;
- The selected sensors of the CSL17 dataset were proven to be practicable for HAR and have good cooperation among each other;
- The single-state HMM modeling of human activity works for the HAR system of the small pilot seven-activity dataset.

The experiments for single-HMM modeling on our comprehensive CSL19 dataset and the external UMS9 dataset will be demonstrated together with the *Motion Units* experiments in Section 5.6.1 to facilitate the comparison between different topologies.

5.1.3 Real-Time HAR Recognizer and Its On-the-Fly Add-On

A wearable real-time HAR system *Activity Signal Kit Echtzeit-Decoder* (ASKED) (Liu and Schultz, 2019) was further implemented based on the experimental results in Section 5.1.2, which verifies the data recording, feature extraction, training, and recognition functionality in the ASK baseline software (see Section 2.3.1) using the CSL17 dataset.

Balance of accuracy versus speed was first studied to improve real-time recognition performance. A shorter step-size of windows shift results in a

shorter delay of the recognition outcomes, but the interim recognition results may fluctuate due to temporary search errors. On the other hand, longer delay due to long step-sizes contradicts a real-time system's characteristics, though it generates more accurate interim recognition results. According to experimental results and user experience, a balancing setting of 400 ms window length with 200 ms window overlap performs the best. These values gave satisfactory recognition results with a barely noticeable delay within 1 second.

After model training, the real-time HAR system starts recording data continuously from the sensors integrated into the knee bandage (see Figure 1.1) and steadily outputs the recognition results with the visualization of the biosignals. The recorded data are displayed serially on the left-hand side of the interface display, and the Top-N (usually set to 3) recognition results in terms of activity classes associated with the calculated probabilities indicated by the length of bars are shown on the right-hand side of the interface display (see Figure 5.6).

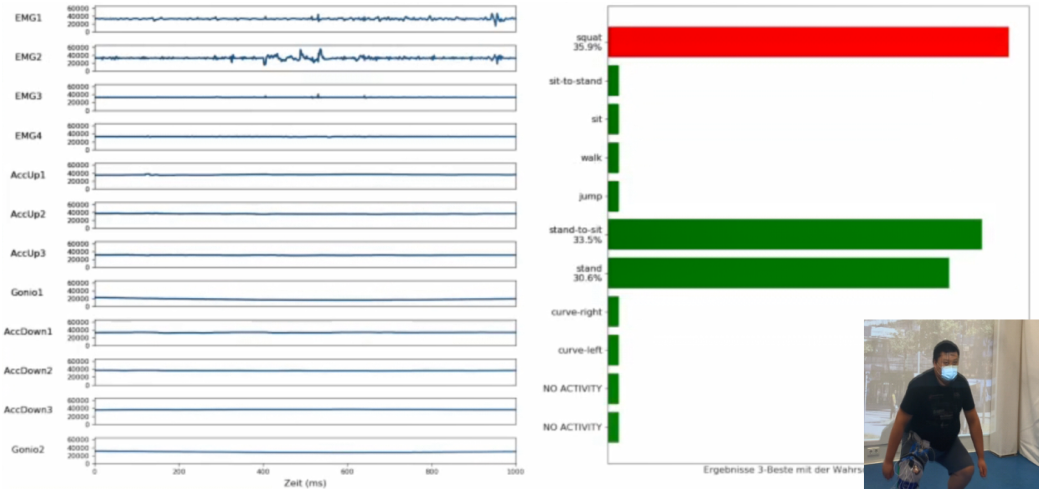


Figure 5.6 – Screenshot of the real-time HAR recognizer ASKED with the on-the-fly added activity "squat." The video in the lower-right corner was recorded synchronously with a mobile phone for an online demonstration.

The GUI menu allows switching biosignals and activities on and off for real-time HAR (see Figure 5.7). This way, it is very straightforward to quickly test the sensors and system properties during development and research.

After successfully testing the real-time HAR system ASKED, an on-the-fly extension named "plug-and-play" (Liu and Schultz, 2019) was implemented, so

the new version of the application is called *Activity Signal Kit Plug-And-PlAy* (ASKPAPA) (see Figure 5.7).

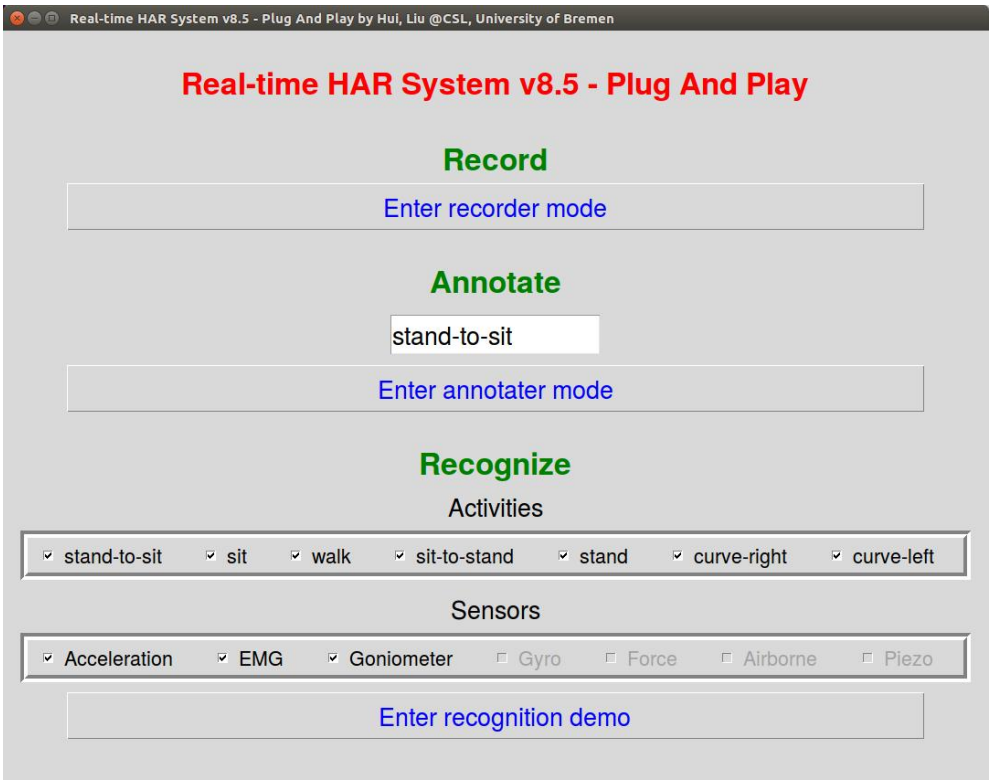


Figure 5.7 – Screenshot of the sensor and activity selection menu of ASKED (Liu and Schultz, 2019).

This add-on can be understood literally: it can load new sensor data on-the-fly, retrain the activity models, and restart the recognition process automatically with the updated activity models. The "plug-and-play" function has the following two use cases:

Providing more training data for an existing activity. We can use the "segmentation and annotation" mode in the ASK baseline software to record more data, such as "walk," generating annotation labels on them at the same time. These new data will automatically serve for training when we restart the real-time HAR system. That is to say, we "plug" more data and "play" with an improved recognition system.

Increase the activity classes to be recognized. The recognition of new activities is enabled easily. By typing a new activity name, such as "squat," in the text-box (see Figure 5.7), we re-run the "segmentation

and annotation" mode, record and label a minimum of 12 instances of "squat." These steps take about three minutes. When finished, the real-time HAR system is started with the new activity "squat" already prepared to be recognized, as shown in Figure 5.6. That is to say, we "plug" data of a new activity and "play" with an upgraded recognizer.

5.1.4 Limitation of Single-State HMM Modeling

To easily illustrate the limitations of single-HMM human activity modeling, let us consider an analogical example in HMM-based speech recognition. Supposing "words" are modeled as the smallest recognizable units, the accuracy on a small dataset might be higher than a system relying on phonemes. However, this kind of recognizer has low training efficiency, low expandability, and low adaptability, which cannot fit a massive vocabulary in a generalized application scenario.

The same holds for HAR, although the performance of applying one activity as one fundamental recognition unit works well with a small dataset in our HMM-based research mentioned above.

5.2 Fixed-Number-of-State HMM Modeling and Experiments

Section 5.1.4 has pointed out that the single-HMM modeling is not optimal, especially when the number of activities to be recognized increases significantly. Therefore, expanding the number of HMMs for each activity is the next important research topic. First, we try to design each activity with the same number (greater than one) of HMMs.

5.2.1 Topology

In (Rebelo et al., 2013), researchers used ten states for each activity and did not mention the reason. We speculate that preferable results determined the number of ten during repeated experiments.

Our HMM-based HAR research has also involved primitively expanding the number of states: (Liu and Schultz, 2019) demonstrated the results of repeated experiments on the different number of states for each activity model, of which the experimental results will be illustrated in Section 5.2.2. (Hartmann et al., 2020) applied six states on all activities for a feature dimensionality study, based on the best performance of repeated experiments, which have

been introduced in Section 4.1. Figure 5.8 illustrates an example of six-state HMM modeling for the activity "walk."

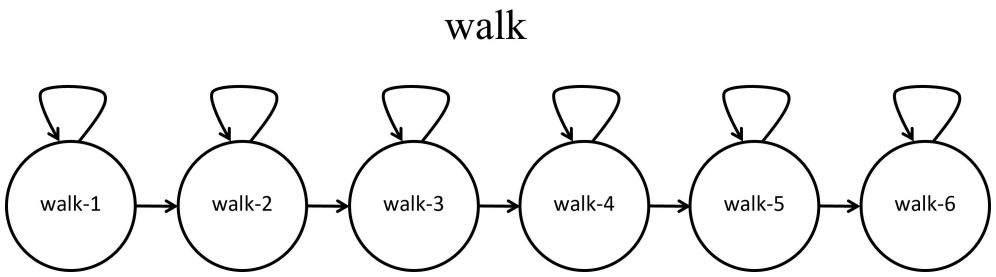


Figure 5.8 – Six-state HMMs of the activity "walk."

5.2.2 Experiments on the CSL18 Dataset

The experiments on the CSL18 dataset aim at the following goals, in the main:

- To verify whether the implementation of the HMM-based HAR research environment in the ASK software is also suitable for more activities by expanding from 7 activities of the CSL17 dataset to 18 activities of the CSL18 dataset;
- To study the increment of each activity's HMM states (from one to a fixed number greater than one) and each state's Gaussians;
- To investigate feature dimensionality and sensor selection by utilizing the larger data volume of CSL18 and the improved modeling from the two goals mentioned above (see Section 4.1).

In the experiments on the CSL18 dataset, each activity consists of a fixed number of HMM states, where a mixture of Gaussians models each state's emission probability. We iteratively optimized the number of HMM states. Before applying the *Split-and-Merge* algorithm for the adaptive determination of the optimal number of Gaussians per state, we also conducted iterative experiments on the fixed number of Gaussians per state. Figures 5.9 and 5.10 demonstrate examples of tuning these two parameters in cross-validation experiments with the setup of 10 ms window length, 5 ms overlap length, and 21-dimensional normalized feature vectors.

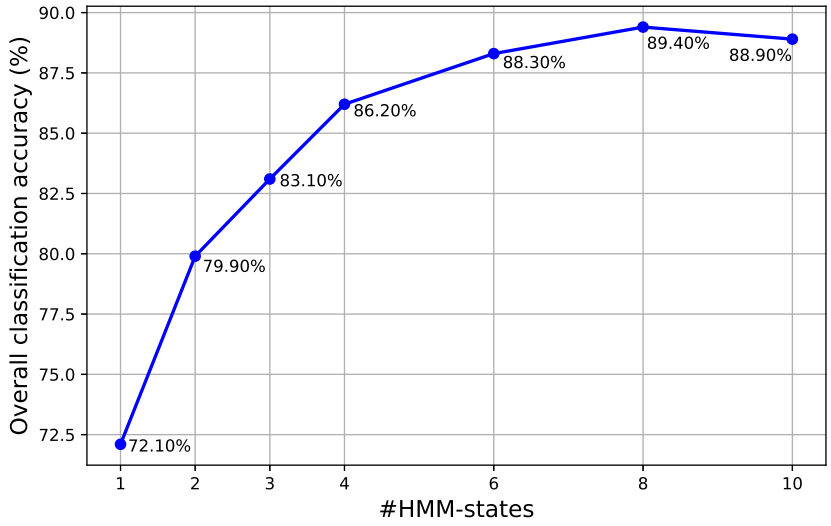


Figure 5.9 – Parameter tuning experiments on the CSL18 dataset: the number of HMM states per activity. Window length: 10 ms; overlap length: 5 ms; dimension of normalized feature vectors: 21; the number of Gaussians per HMM state: 5. (Liu and Schultz, 2019).

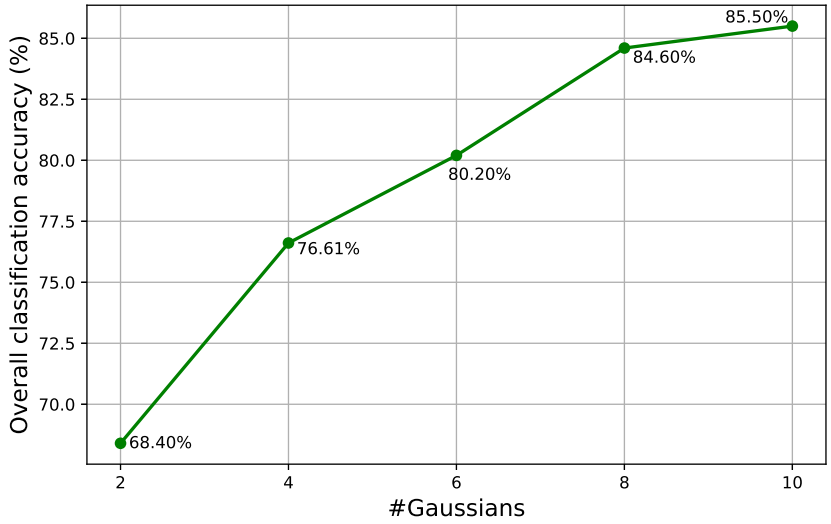


Figure 5.10 – Parameter tuning experiments on the CSL18 dataset: the number of Gaussians per state. Window length: 10 ms; overlap length: 5 ms; dimension of normalized feature vectors: 21; the number of HMM states per activity: 2 (Liu and Schultz, 2019).

Due to the limited data quantity, we stopped evaluating the number of Gaussians at 10 to achieve reliable results. Similar experiments for tuning different parameters were carried out thoroughly, and we concluded that the application of eight HMM states and ten Gaussians offers the best recognition results for the CSL18 dataset without applying the *Split-and-Merge* method.

By using these current best-performing parameters, the overall person-dependent recognition accuracy achieves almost 90%. Figure 5.11 illustrates the recognition results in percentage from one cross-validation experiment in the confusion matrix. Table 6 gives the criteria precision, recall, and F-Score in each activity's average from cross-validation experiments.

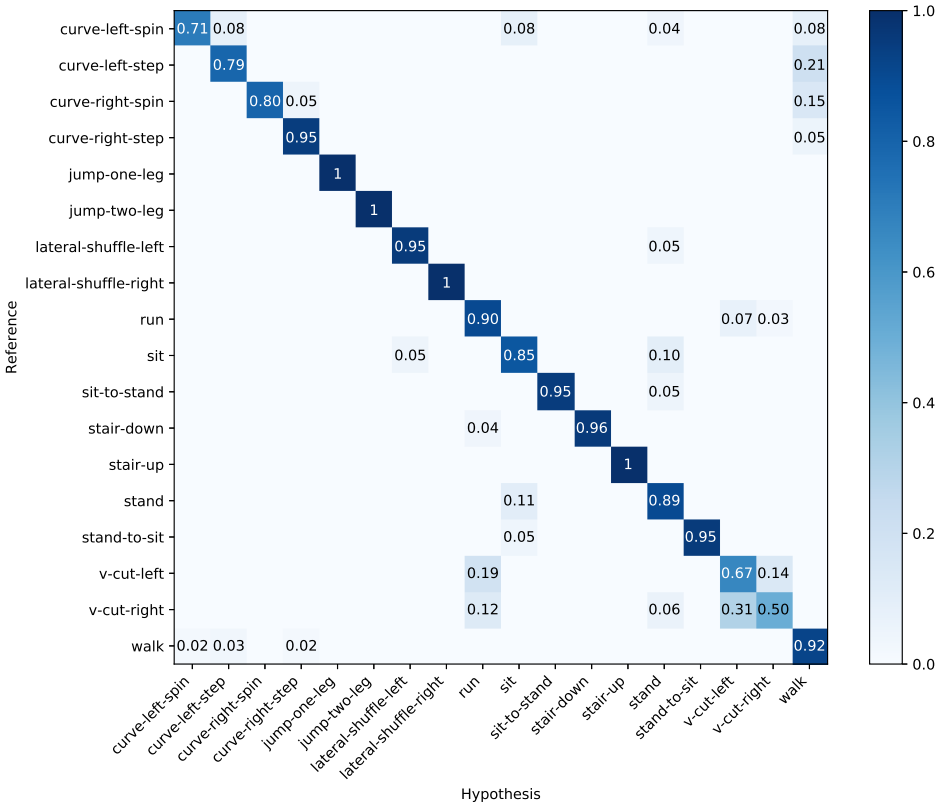


Figure 5.11 – Confusion matrix of recognition results in percentage from one cross-validation experiment on the CSL18 dataset (Liu and Schultz, 2019).

As can be seen in Figure 5.11 and Table 5.3, activities "jump-one-leg," "jump-two-leg," "shuffle-right," and "walk-upstairs" were correctly recognized in each experiment.

Table 5.3 – Criteria in each activity's average from cross-validation experiments on the CSL18 dataset (Liu and Schultz, 2019).

Activity	Precision	Recall	F-Score
sit	0.85	0.78	0.81
stand	0.90	0.75	0.81
sit-to-stand	0.95	1.00	0.97
stand-to-sit	0.96	1.00	0.98
walk-upstairs	1.00	1.00	1.00
walk-downstairs	0.96	1.00	0.98
walk	0.92	0.88	0.90
walk-curve-left	0.79	0.83	0.80
spin-left	0.71	0.91	0.78
walk-curve-right	0.95	0.88	0.91
spin-right	0.80	1.00	0.89
run	0.90	0.86	0.87
V-cut-left	0.66	0.70	0.64
V-cut-right	0.50	0.56	0.50
shuffle-left	0.96	0.96	0.95
shuffle-right	1.00	1.00	1.00
jump-one-leg	1.00	1.00	1.00
jump-two-leg	1.00	1.00	1.00
Global accuracy: 0.90			

It should be pointed out that the above-described recognition experiments on the CSL18 dataset are based on the seven-subject complete data collection. Later, we found that three subjects' data was incomplete due to the loss of sensor signals during the collection. The above experimental results using the seven-subject data are biased due to incomplete data but still have good reference value to a certain extent, as we can regard it as a simulation of the real-world data recording with uncertainty. In real application scenarios, we cannot guarantee that the HAR system always obtains "intact" sensor data due to transmission obstruction, occasional sensor instability, among others.

The current official version of the CSL18 dataset only contains the remaining four subjects' data (see Section 3.4). Except for the previous research on feature dimensionality (see Section 4.1), we did not further conduct training-recognition and activity modeling research on the reduced four-subject CSL18 dataset because we have adopted the more extensive and comprehensive CSL19 dataset of verified correctness.

The experiments for fixed-number-of-state HMM modeling on the CSL19 and the UMS9 datasets will be demonstrated together with the *Motion Units* experiments in Section 5.6.2 to facilitate the comparison between different topologies.

5.2.3 Limitation of Fixed-Number-of-State HMM Modeling

No matter the fixed number of states, each state's meaning is still unknown, similar to a DNN.

Hence, there are two problems worth further exploring:

- 1. Could/should each activity contain a separate, explanatory number of states?

If the answer is "yes," the explosion of possible combinations renders seeking a model based on repeated experiments unfeasible. Therefore:

- 2. Is there an approach to design HMMs of human activities more rule-based, normalized over blindly "trying"?

As follows, we attempt to solve the two problems by proposing a novel modeling technology.

5.3 Phase and State Partitioning Modeling of Human Activities

In order to solve the two questions listed in Section 5.2.3 and to illustrate the idea of human activity modeling more clearly, we take a typical HMM-based word modeling topology in speech recognition as an example.

5.3.1 Inspiration from Speech Recognition

Figure 5.12 shows a three-state Bakis-model (Bakis, 1976) constructing each phoneme (/d/ and /ɪ/). Each state, also called sub-phoneme, models parts of a phoneme (begin/middle/end). Following this topology, we can practically build a pronunciation dictionary by concatenating phonemes without regard to contextual dependency for simplifying notations, as shown in Table 5.4.

Phoneme and sub-phoneme are traditionally used in *Automatic Speech Recognition* (ASR). How to analogically define the topology of human activity?

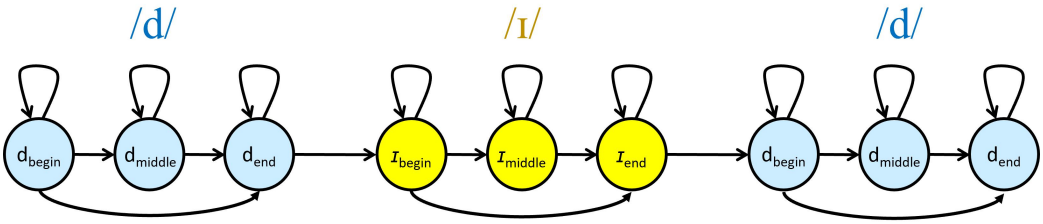


Figure 5.12 – A typical linear left-right HMM of the phoneme sequence "did" (Liu et al., 2021).

Table 5.4 – A simple pronunciation dictionary.

Word	Dictionary (sequence of phoneme)
did	/d/ — /ɪ/ — /d/
dig	/d/ — /ɪ/ — /g/
gid	/g/ — /ɪ — /d/
gig	/g/ — /ɪ — /g/
digged	/d/ — /ɪ/ — /g/ — /d/
gigged	/g/ — /ɪ/ — /g/ — /d/
...	...

5.3.2 Gait Analysis for Phase and State Partitioning

Some activity recognition systems only focus on human activities happening in specific body parts, e.g., hands, such as Airwriting (Amma et al., 2010). However, it is not surprising that most sensor-based HAR research for ambulation activities uses and sometimes only uses body-worn sensors placed below the waist, such as pant's pockets, thighs, knees, shanks, and feet because the lower body plays a decisive role in position translation, an essential part of most human activities. This brings us to the research on gait analysis, which provides us the first clue for studying activity partitioning.

In (Whittle, 1996) and (Whittle, 2014), the author commonly distinguishes two phases into seven sub-phases during one full gait cycle, where each sub-phase has its corresponding physical quantities' characteristics and typical activation of major muscle groups, of which the sagittal joint angles of different body parts are demonstrated in Figure 5.13.

Besides, the initial contact event is often used as the start/end event of a gait cycle and may explicitly be modeled as seen in (Arous et al., 2018) and (Mezghani et al., 2013), bringing the total number of sub-phases to eight.

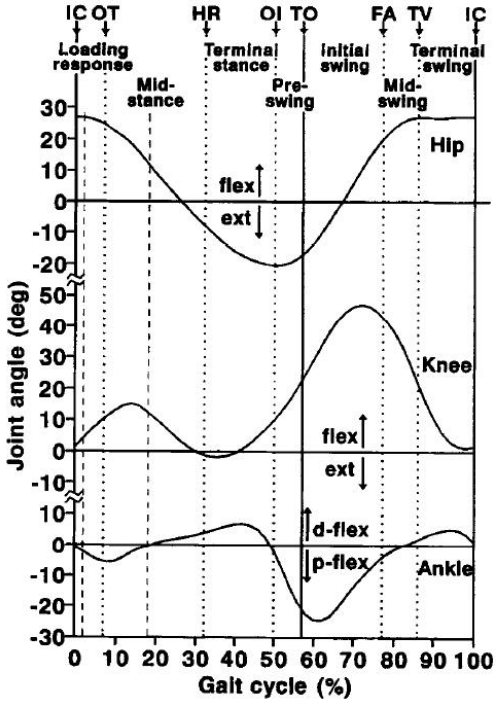


Figure 5.13 – Sagittal joint angles of hip, knee, and ankle during a single gait cycle by a normal subject. IC: initial contact; OT: opposite toe off; HR: heel rise; OI: opposite initial contact; TO: toe off; FA: feet adjacent; TV: tibia vertical (Whittle, 1996).

However, this does not mean that directly applying "seven" or "eight" as the number of states to all activities will achieve the best accuracy, as clarified in (Hartmann et al., 2020). For example, some of these events, such as "pro-swing" and "mid-swing," have a too short duration that does not fit a single HMM state. There is still a gap to bridge from biological research and sports science to informatics modeling.

In collaboration with kinesiologists of the *Institute of Sport and Sports Science* at KIT, we decided to model one gait cycle as five states, representing three and two states respectively in the "two-phase" gait analysis, based on the investigation of sports and gait science knowledge (e.g., (Whittle, 1996), (Whittle, 2014), (Arous et al., 2018), and (Mezghani et al., 2013)), the phase duration analysis with the forced alignment visualization (e.g., Figure 5.15), and the inspiration from speech recognition (e.g., Figure 5.12). Numerous experiments using different numbers of states for gait-based activities, such as those introduced in Chapter 4 and Section 5.2, also have verified the superior performance of this division method.

Figure 5.14 depicts the modeling scheme of a typical gait-based activity "walk." A complete gait is divided into two phases as observed from one leg: the stance (ground-contacting) phase and the swing phase. In the stance phase, we adopted three states (initial/middle/terminal) by analogizing the sub-phoneme in speech recognition, while in the swing phase, we designed only two states (initial/terminal). These states can go by the name of **sub-phases**.

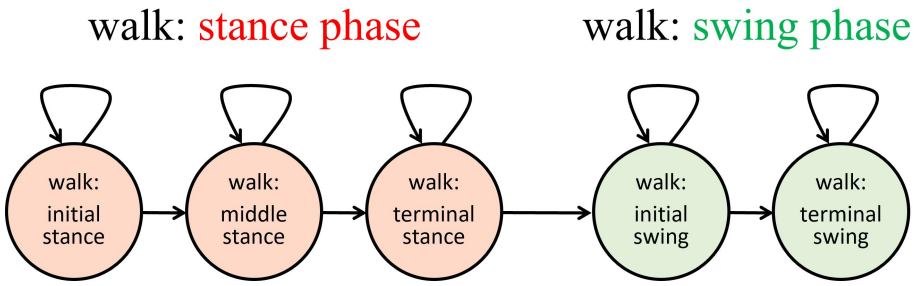


Figure 5.14 – A linear left-right HMM for one gait in the activity "walk" based on phase and state partitioning. Red: states/sub-phases in the stance phase; green: states/sub-phases in the swing phase (Liu et al., 2021).

Figure 5.15 gives an example of the HMM forced alignment experiments illustrating that the HMM learns correct meanings, as the states, such as "initial stance" and "middle stance," are correctly and automatically labeled according to the sagittal knee joint angle changing over time in Figure 5.13 (the value axis is reversed), which primitively verified the phase and state partitioning of gait-based activities.

Based on the phase duration statistics and visualization (e.g., Figure 5.15), We have also investigated that during a regular "walk" activity, the duration of the stance phase varies between 200 ms and 800 ms, and the swing phase between 200 ms and 600 ms, which provides an essential reference for the window length selection in subsequent tasks of training and recognition.

Other gait-based activities, such as "run," "walk-upstairs/downstairs," "shuffle-left/right," and "V-cut-left/right," follow closely to the pattern of "two-phase-five-state" gait modeling, as these activities can also be divided into several gaits, where each gait's knee joint angle changing is also in accordance with Figure 5.13. However, they have a difference in speed, intensity, or direction compared with a regular "walk" activity. The forced alignment visualization results of these activities are very similar to Figure 5.15. Also, this partitioning result's performance has been verified through Numerous experiments using different numbers of states.

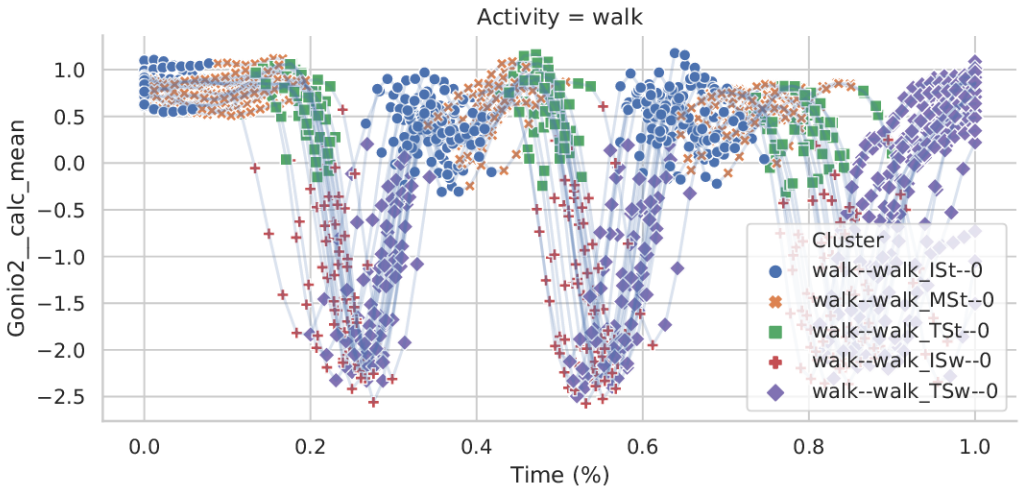


Figure 5.15 – HMM forced alignment based on the phase and state partitioning of the activity "walk" in the CSL19 dataset. 20 randomly sampled sequences are plotted on a percentage-based time axis. The lines show each sequence, while colors/shapes denote each vector's state as determined by the alignment (Liu et al., 2021).

Figure 5.16 gives an analogical modeling example of the activity category "V-cut."

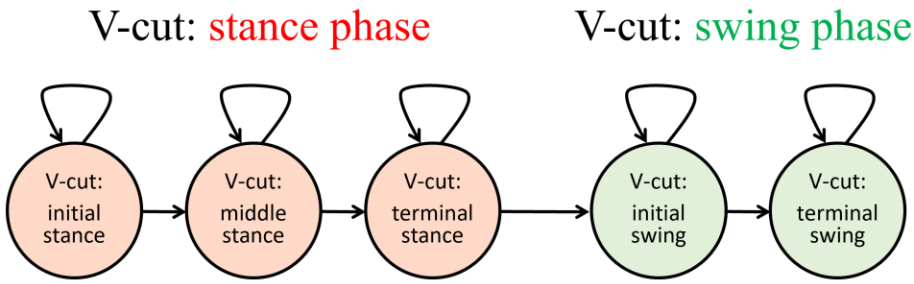


Figure 5.16 – A linear left-right HMM for the activity category "V-cut" based on phase and state partitioning. Red: states/sub-phases in the stance phase; green: states/sub-phases in the swing phase.

In the study of partitioning, model generalization (see Section 5.5) has not been applied yet, so currently, in the entire HMM-dictionary, no activities share states.

5.3.3 Partitioning of Non-Gait-Based Activities

We continued to design the HMM modeling of more activities progressively through the experience accumulation from the modeling procedure of gait-based activities. The "two-phase-five-state" topology fits mainly the gait-based activities. Consequently, we must investigate the sports knowledge for each new activity and analyze the data to derive the quantities and topology.

Static activities like "stand" and "sit" can be described using only one state (see Figure 5.17), which can be understood from the perspective of their relatively stable sensor signals and confirmed by numerous previous experiments, such as those introduced in Chapter 4 and Section 5.1.2.

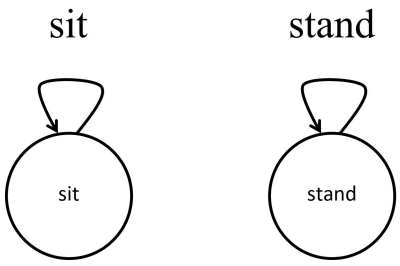


Figure 5.17 – HMMs for the one-state activities "sit" and "stand."

The single-state HMM modeling of the in-vertical-shifting activities where the feet stay in place, like "sit-to-stand" (stand up) or "stand-to-sit" (sit down) achieved good recognition accuracy on the CSL19 dataset (97.2% for "sit-to-stand" and 99.2% for "stand-to-sit"; see Figure 5.25). Repeated experiments showed that the use of dual-state HMMs (see Figure 5.18) could improve the results (99.0% for "sit-to-stand" and 100% for "stand-to-sit"; see Figure 5.29). Further increasing the number of states does not benefit the recognition rate, but increases the training cost, as the analysis in Section 5.6.2 indicates.

Also in collaboration with kinesiologists of the Institute of Sport and Sports Science at KIT, "jump" is divided into three phases (takeoff/shift-up/land) with five sub-phases (see Figure 5.19), based on the duration analysis and repeated experiments. This partitioning has a certain degree of physical interpretability from the perspective of sports kinematics, but it may not always provide the best recognition rate: it works much better than the single-state topology but slightly worse than the fixed-six-state topology on the CSL19 dataset; on the UMS dataset, it takes the lead (see Sections 5.6.1—5.6.3). In the future, we will conduct a more in-depth study on jumping activities.

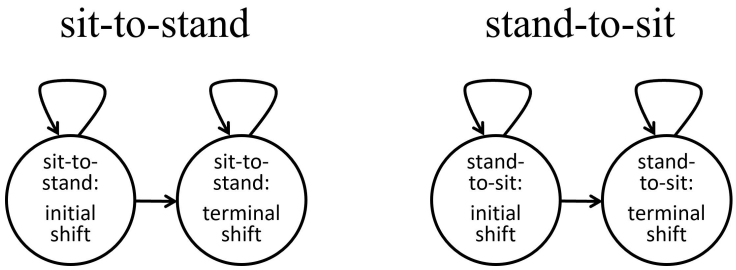


Figure 5.18 – Linear left-right HMMs for the dual-state activities "sit-to-stand" and "stand-to-sit" based on phase and state partitioning.

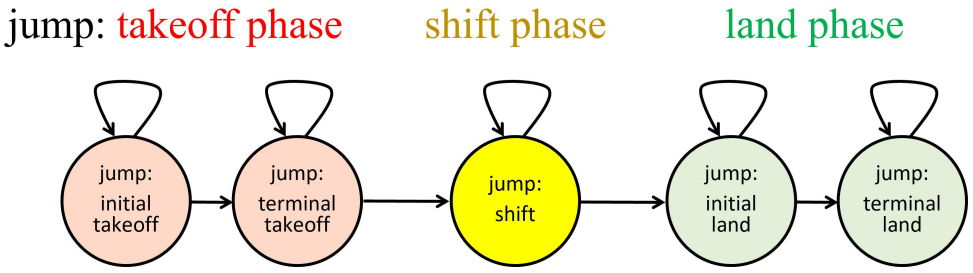


Figure 5.19 – A linear left-right HMM for the three-phase-five-state activity "jump" based on phase and state partitioning. Red: states/sub-phases in the takeoff phase; yellow: state/sub-phase in the shift phase; green: states/sub-phases in the land phase.

5.3.4 Denotation of Phase and State Partitioning of Human Activities

In the following, we use **I**, **M**, and **T** to denote the earlier mentioned sub-phases "initial," "middle," and "terminal," respectively, while the phase names like "stance," "swing," "shift," "takeoff," and "land" introduced above are abbreviated as **St**, **Sw**, **Sh**, **To**, and **La**, respectively. For example, **TSw** represents the state "terminal swing." Table 5.5 lists the denotations applied in the following sections' research with their abbreviation.

5.3.5 Partitioning Topology on the CSL19 and the UMS9 Datasets

Tables 5.6 and 5.7 list the latest versions of the phase and state partitioning results on the nine-ADL-subset of the external UMS dataset (UMS9) and the CSL19 dataset, based on the most reasonable performance of parameter tuning experiments so far.

Table 5.5 – Some states/sub-phases of universal significance and their abbreviation.

State/sub-phase	Abbreviation
Initial Stance	ISt
Middle Stance	MSt
Terminal Stance	TSt
Initial Swing	ISw
Terminal Swing	TSw
Initial Takeoff	ITO
Terminal Takeoff	TTO
Shift	Sh
Initial Shift	ISh
Middle Shift	MSh
Terminal Shift	TSh
Initial Land	ILa
Terminal Land	TLa

Table 5.6 – The number of HMM states for each activity based on the latest partitioning practice on the nine-ADL-subset of the UMS dataset (UMS9).

Activity	Number of HMM states
StandingUpFromSitting	2
SittingDown	2
StandingUpFromLaying	5
LyingDownFromStanding	5
Walking	5 (per gait)
GoingDownStairs	5 (per gait)
GoingUpStairs	5 (per gait)
Running	5 (per gait)
Jumping	5
Total number of HMM states	**39**

Table 5.7 – The number of HMM states for each activity based on the latest partitioning practice on the CSL19 dataset.

Activity	Number of HMM states
sit	1
stand	1
sit-to-stand	2
stand-to-sit	2
walk, including walk-curve-left (90°) and walk-curve-right (90°)	5 (per gait)
walk-upstairs	5 (per gait)
walk-downstairs	5 (per gait)
spin-left-left-first	5
spin-left-right-first	5
spin-right-left-first	5
spin-right-right-first	5
run	5 (per gait)
V-cut-left-left-first	5 (per gait)
V-cut-left-right-first	5 (per gait)
V-cut-right-left-first	5 (per gait)
V-cut-right-right-first	5 (per gait)
shuffle-left	5 (per gait)
shuffle-right	5 (per gait)
jump-one-leg	5
jump-two-leg	5
Total number of HMM states	**86**

The two datasets contain most of the main ambulation activities, and the sensors applied are mainly placed in the lower body. In the current partitioning and the subsequent generalization experiments, "walk-curve-left (90°)" and "walk-curve-right (90°)" are merged into "walk" because numerous experimental results so far show that it is hard to distinguish between these three activities. From the perspective of horizontal comparison, we have not found any other datasets or literature at present, which distinguish the "in-progress small-angle turn while walking" as an independent activity. Therefore, we have merged "walk," "walk-curve-left (90°)," and "walk-curve-right (90°)" in the latest version in order to avoid such obstacles for proceeding with the following modeling research smoothly. However, it does not imply that

this will always be the case in the future. We plan to research on robust recognition between these three activities.

It is worth mentioning that the number of generated states based on sports science knowledge and signal duration analysis is not necessarily the final optimal solution; thereupon, a certain amount of repeated experiments for fine-tuning may be required. Compared with the blind "search" or the use of a fixed number of states for all activities uniformly, the phase and state partitioning is more interpretable and expandable, which serves as a benchmark model for the following model generalization research.

The experimental results of the partitioning modeling on the CSL19 and the UMS9 datasets will be demonstrated together with the *Motion Units* experiments in Section 5.6.3 to facilitate the comparison between different topologies. Before that, we will use the partitioning topology to conduct feature selection research, hoping to obtain an ideal feature combination for subsequent model generalization (*Motion Units*) research and performance comparison between models.

5.4 Feature Selection Experiments Based on Partitioning Modeling

As emphasized in (Domingos, 2012), "*At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used.*" The feature selection will affect the model training of HAR research and the quality of the final system.

After having a partitioning model of the CSL19 and the UMS9 datasets according to the method described in Section 5.3.5 and verifying its applicability compared to the single-state and fixed-number-of-state models through iterative experiments, we carried out feature selection experiments on this basis to obtain an ideal feature combination for a larger number of subsequent model generalization (*Motion Units*) experiments.

For each channel of each sensor, there are dozens of selectable features. Therefore, feature selection will face two problems:

- How can we discover as many appropriate features as possible and compute them efficiently?
- Which method(s) should be used to compare and select between these massive features?

The following two sections address these two aspects respectively.

5.4.1 Time Series Feature Extraction Library (TSFEL)

In order to save research time and introduce as many features as possible, we apply an open-source *Python* package named *Time Series Feature Extraction Library* (TSFEL), which provides support for fast exploratory analysis supported by an automated process of feature extraction on multidimensional time series. As a cooperative development member, the author of this dissertation has also contributed to this open library and shared release publication (Barandas et al., 2020).

Figure 5.20 describes the TSFEL processing pipeline. It takes time series (arrays previously loaded in memory or stored in files) as inputs for the extraction processor. As an essential step for handling multidimensional time series, a set of preprocessing methods, such as those introduced in Session 2.5.2, is afterward applied to ensure signal quality and time series synchronization for an appropriate window calculation process. The resulted features are saved using a standard schema, of which each line corresponds to a window of the extracted features along the corresponding columns. Nearly 60 different features can be extracted across temporal, statistical and spectral domains.

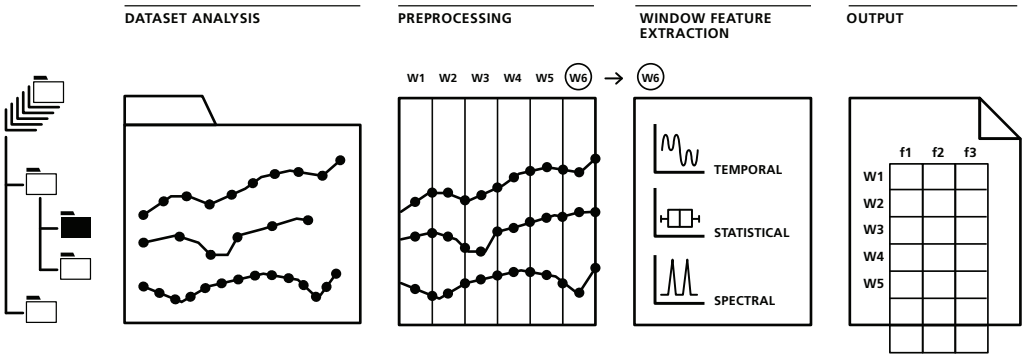


Figure 5.20 – TSFEL pipeline: dataset analysis, signal preprocessing, feature extraction, and output (Barandas et al., 2020).

Figure 5.21 shows an illustrative example with a typical pipeline to extract features in the context of HAR applying a subset of the CSL19 dataset composed of time series retrieved by the accelerometer, gyroscope, and electrogoniometer from two subjects performing four activities: "stand," "sit," "stand-to-sit," and "sit-to-stand." Five exemplar features extracted in such a pipeline are illustrated in a horizon plot in Figure 5.21.

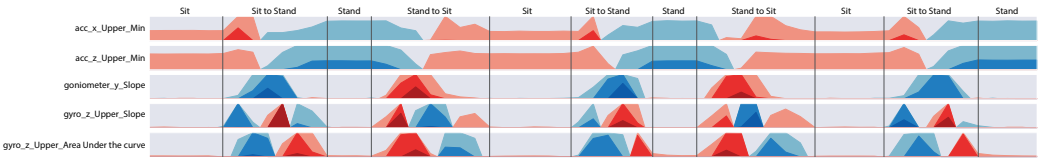


Figure 5.21 – Horizon plot representation from five exemplar features. The vertical lines correspond to ground truth annotations to discriminate between different classes using the protocol-for-pushbutton mechanism (Barandas et al., 2020).

5.4.2 Approaches for Feature Selection and Experiments on the CSL19 Dataset

We selected 35 features from TSFEL, as listed in Table 5.8, instead of using all the features in the library because the calculation cost of the remaining features is very high, which will increase each feature selection's running time by several days, even up to the level of weeks and months. It is also foreseeable that one of our research goals — a user-oriented real-time HAR system, will not favor these features with high computation time.

Table 5.8 – Features chosen from TSFEL for feature selection experiments. diff.: difference; dev.: deviation; Max.: Maximum (Hartmann, 2020).

Temporal domain	Statistical domain	Frequency domain
Absolute Energy	Maximum	FFT mean coefficients
Area under the curve	Mean	Max. frequency
Autocorrelation	Minimum	Max. power spectrum
Centroid	Standard dev.	Spectral centroid
Distance	Interquartile range	Spectral decrease
Mean absolute diff.	Kurtosis	Spectral distance
Mean difference	Mean absolute dev.	Spectral kurtosis
Sum of absolute diff.	Root mean square	Spectral roll-off
Total energy	Skewness	Spectral roll-on
Slope		Spectral skewness
Zero crossings		Spectral slope
		Spectral spread
		Spectral variation
		Fundamental frequency
		Power bandwidth

The base feature implementation was taken from TSFEL and re-implemented applying NumPy (NumPy, 2021) vectorization, achieving an average 22 times

speed enhancement on the CSL19 dataset. The performance comparison between the vectorized and library features can be found in (Hartmann, 2020). The vectorized features were confirmed to run correctly with the tests provided by TSFEL.

The following describes the results of the most informative feature selection methods, the greedy forward selection based on the HMM recognition experiments. In the greedy forward selection, each step involves a complete incremental HAR model training and recognition experiment on each accumulated set of candidate features. Therefore, it consumes a considerable amount of time, but its selection results produce an improved reference value. One complete greedy forward selection experiment took about twelve days on 48 cores at 2.5 GHz CPU cores. Figures 5.22 and 5.23 demonstrate the top 25 greedy feature selection results of two representative experiments with different window and overlap length settings.

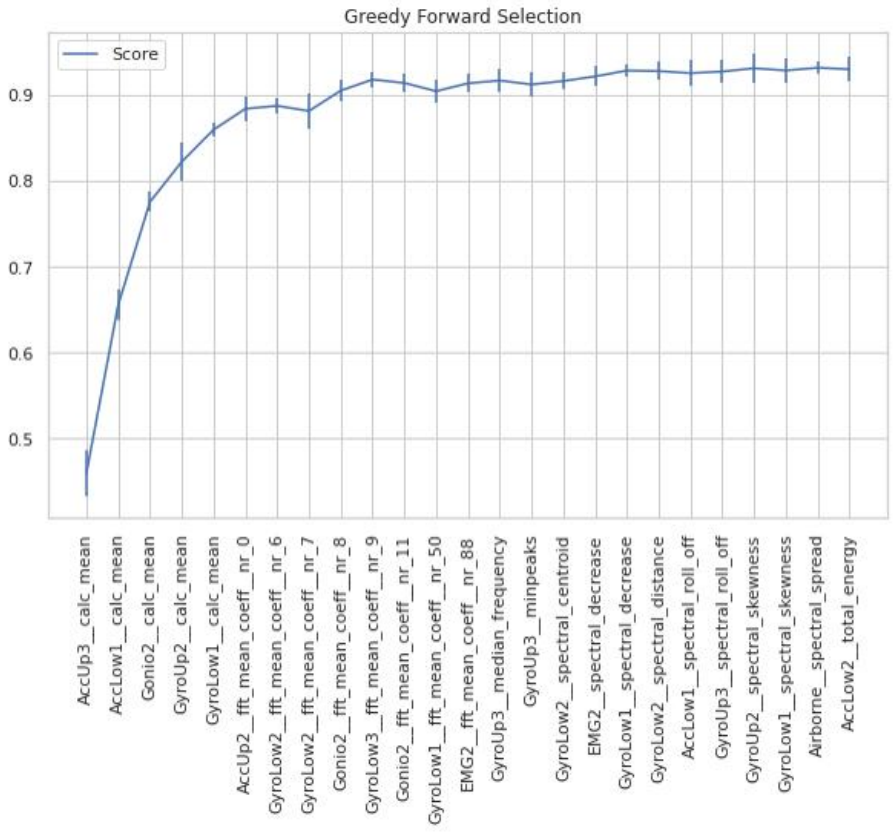


Figure 5.22 – The top 25 greedy forward selection results based on the features extracted by TSFEL on the CSL19 dataset using 150 ms of window length and 20% of overlap length.

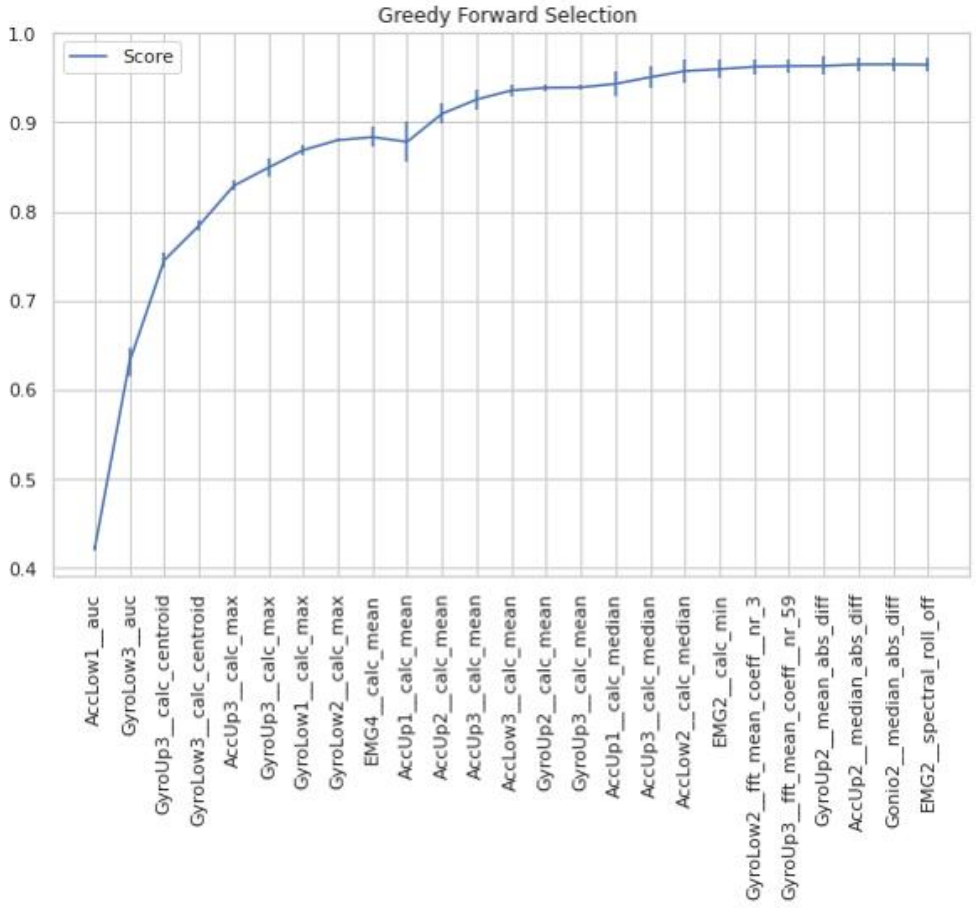


Figure 5.23 – The top 25 greedy forward selection results based on the features extracted by TSFEL on the CSL19 dataset using 200 ms of window length and 50% of overlap length.

Both Figures 5.22 and 5.23 exhibit that the recognition rate is very low when there are only a few selected features initially. At this time, each additional feature improves the recognition rate significantly. When selecting the 8th to 10th features, the two sets of settings have similarly experienced a process of decreasing recognition rate, i.e., the current best additional features cannot further optimize the training and decoding. The recognition rates afterward exhibited fluctuations (Figure 5.22) or a steady ascending trend (Figure 5.23). This tendency provides us with the experience that we cannot use the "first falling point" as the terminal of the greedy forward feature selection. Instead, we used the entire feature set for feature selection which ran twelve days, as mentioned above.

5.4.3 Results of Feature Selection on the CSL19 and the UMS Datasets

According to the results of the greedy forward feature selection on the TSFEL features, the recognition of the CSL19 dataset could not be further improved and retained the original features and 34-dimensional feature space using *average* and RMS, as introduced in Section 4.2.2.

The same selection procedure on the full UMS dataset (also containing falls) yielded a 20-dimensional combination of time and frequency features, including, for instance, min, slope, and spectral spread (Liu et al., 2021).

5.5 Model Generalization and Motion Units (MUs)

The next step is to study the generalizability of the states obtained from phase and state partitioning to simplify the overall modeling further, for which we also start with the analogy of speech recognition.

5.5.1 Inspiration from Speech Recognition

In speech recognition, this problem of generalization has already been appropriately and adequately solved. Phonemes (such as /d/, /ɪ/, and /g/ in Table 5.4), whose amount is very small, are used as base units in speech recognition to represent all words in the language as sequences. Thus, the training of the models and the utilization of the data corpus becomes quite efficient because data for training the models can be shared. For example, the speech data of "gid" and "dig" are used together to train the models of "d," "g," and "i."

Even taking account of contextual dependency and co-articulation, each language has its typical set of primarily fixed phonemes, which is the key to the model generalization of speech recognition.

Based on the inspiration from speech recognition, we want to explore a method/design that can provide generalized activity modeling references for most datasets and application scenarios. Of course, it can also accommodate any exceptional cases and particular circumstances.

5.5.2 Approaches for Generalization

To explain our proposed model generalization scheme concisely, let us take some of the gait-based activities in the CSL19 and the UMS datasets as an example. In non-generalized partitioning modeling, "walk," "walk-upstairs," and "run" all contain two phases and five states (see Tables 5.6 and 5.7) that are different from each other. Not to mention repetitions like multiple gait-cycles in each activity. Which of these states can be generalized?

The most straightforward consideration is from inertial biosignals' general knowledge: the **ISw** states and the **TSw** states in these three activities' swing phases cannot be merged. "Walk" and "walk-upstairs" have different translational directions, while "walk" and "run" have different movement speeds.

Let us concentrate on the generalizability of the remaining three states in the stance phases of these three activities. We started by choosing all possible candidates with which we saw the best opportunities and tried to merge them. Taking the CSL19 dataset as an example, besides inertial sensors, we also use EMG sensors on the thigh and shank, respectively. It can be estimated that as time goes by, the different activities' states will become more and more different. So, we attempted to merge the **ISt** states of "walk" and "walk-upstairs" at the very beginning and confirmed that they are mergeable since experiments show no significant recognition lost compared to the non-generalized models.

Along this trajectory, we continue to conduct several repeated experiments on other states (e.g., **MSt** and **TSt**) and other gait-based activities to obtain a reliable generalization result. Notably, the complexity of the repeated experiments in this step is not large-scaled. The preliminary partitioning and theoretical generalization design have already provided an appropriate baseline model.

Figure 5.24 provides a representative example of repeated generalization experiments. As mentioned in Section 3.5.1, in the CSL19 dataset, we divided the single-gait "spin-left" and "spin-right" activities into two sub-activities, left foot first and right foot first, respectively. Taking "spin-left" as an example, we observed in the data collection sessions that whether the left foot or the right foot is first, the right leg will draw a similar 90° turn. Since all the sensors are placed near the knee of the right leg, can the two sub-activities of "spin-left" share the same swing phase states?

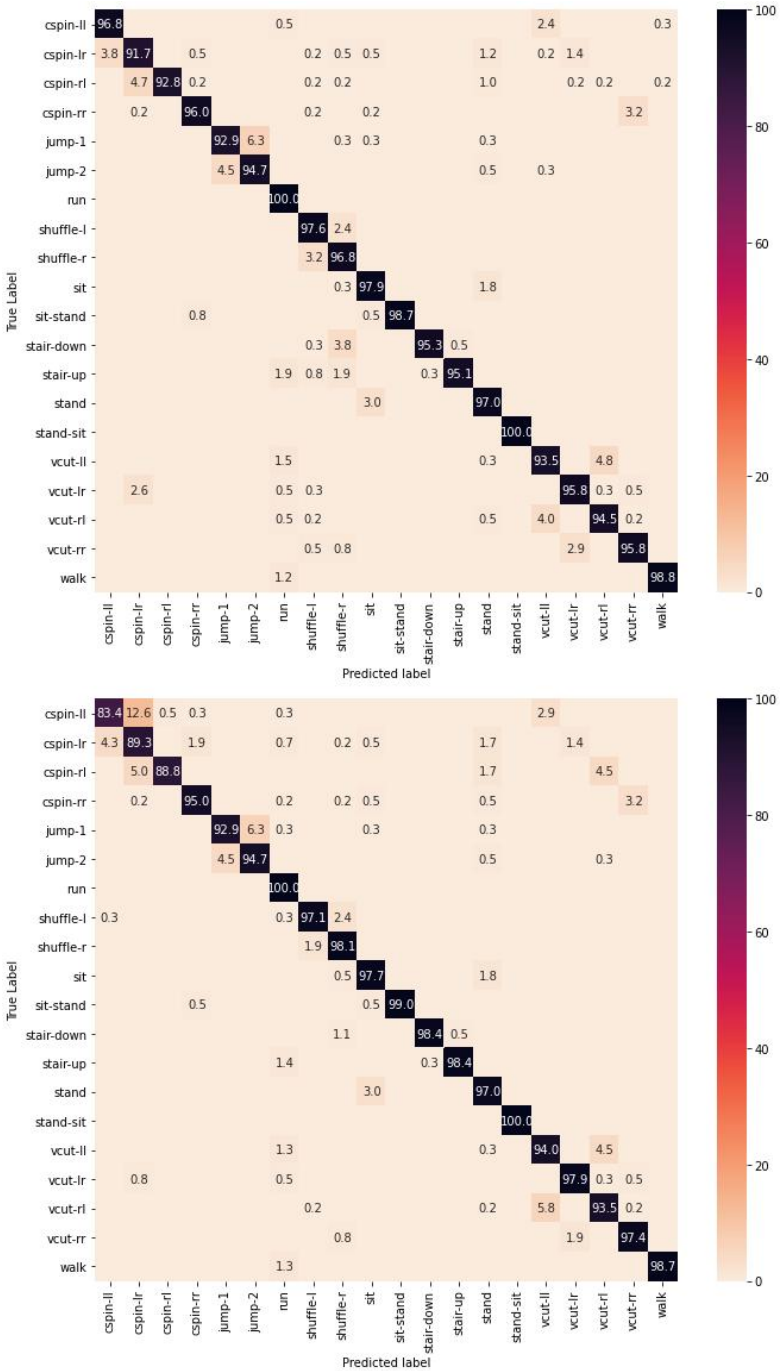


Figure 5.24 – Confusion matrix of cross-validation person-independent recognition results in percentage on the CSL19 dataset: state merging experiments of the swing phases in "spin" activities. Top: no shared states; bottom: "spin-left-left-first" and "spin-left-right-first" shared both **ISw** and **TSw** states in the swing phase, and so as "spin-right-left-first" and "spin-right-right-first."

Figure 5.24 answers the question negatively. We can obtain the following observation from the confusion matrices and other criteria such as the global recognition accuracy:

- After merging, the recognition rate of the activities other than the "spin"s has no or minor changes;
- The recognition performance of the "spin"s has significantly deteriorated;
- The confusion between the "spin" activities has become evident;
- As a result, the global recognition accuracy is also reduced.

Therefore, the experimental results confirm that the "spin" activities' swing phase states cannot be shared. Similar experiments have been repeated on different activities and different states.

For non-gait-based activities, through the above-introduced approaches of repeated experiments, we can also judge whether the candidate states are generalizable or not. For example, the two activities "LyingDownFromStanding" and "SittingDown" in the UMS dataset share the **ISh** state, while the two activities "StandingUpFromLaying" and "StandingUpFromSitting" share the **TSh** state. In contrast, for the CSL19 dataset, no state in the "jump-one-leg" and "jump-two-leg" activities is generalizable, as the state merging experiments objects to all sharing hypotheses.

In Section 5.6.4, the most optimized experimental results of model generalization so far will be demonstrated and compared with other HMM modeling topologies introduced above.

5.5.3 Motion Units

The states of each activity in each dataset are designed based on the merging strategy and repeated experiments described in Section 5.5.2, regardless of whether they are used repeatedly in the applied modeling dictionary. We call these states *Motion Units* (MUs) and they are the generalized recognizable units composing each human activity in the HAR system, analog to phonemes in speech recognition. The same activities (more precisely, activities with the same name) in both CSL19 and UMS datasets are modeled using the same HMM state sequence (MUs).

It is worth noting that although we used the word "generalized," in the experiments described in the next sections, some states only appear in a certain activity and not be reused (see Figures 5.34—5.36). However, this

does not mean that these states cannot be shared. The number of activities in our research's datasets does not support the reuse of some states, suggesting a very different point between HAR and speech recognition research. In fact, if more different activities are defined, collected and modeled, the reusability of MUs will be significantly improved.

5.6 Comprehensive Human Activity Modeling Experiments of 4 Topologies on the CSL19 and the UMS9 Datasets

Numerous experiments comparing four topologies on the CSL19 and the UMS9 datasets are conducted to evaluate the phase and state partitioning and MU-based generalization design's real-world applicability. In each experiment, the HMM model is evaluated using person-independent leave-one-subject-out cross-validation.

The implemented recognizers and chosen hyperparameters are based on Table 4.5 and Section 4.3. The features are determined based on the experimental results in Section 5.4.3. The recognizer for the CSL19 dataset uses 100 ms*Hamming* windows with an overlap of 50 ms and a normalization over the whole sequence. On the UMS dataset, 400 ms *Hamming* windows with an overlap of 320 ms and normalization is used (Hartmann et al., 2021). Gaussian mixture models field the HMM emission model. Due to a low chosen threshold and maximum Gaussians in the *Split-and-Merge* training algorithm, the total number of Gaussians scales proportionally with the number of states in the HMM topology.

5.6.1 Single-State Topology

The lower reference of the partitioning and MU experimental study is created using a single-state topology, which has been proven effective in (Liu and Schultz, 2018), as described in Section 5.1.2.

Figure 5.25 and Table 5.9 depict the confusion matrices and the criteria precision, recall, F-scores, and recognition accuracy of the single-state HMM topology recognition results on the CSL19 dataset, respectively, while Figure 5.26 and Table 5.10 show the results on the UMS9 dataset.

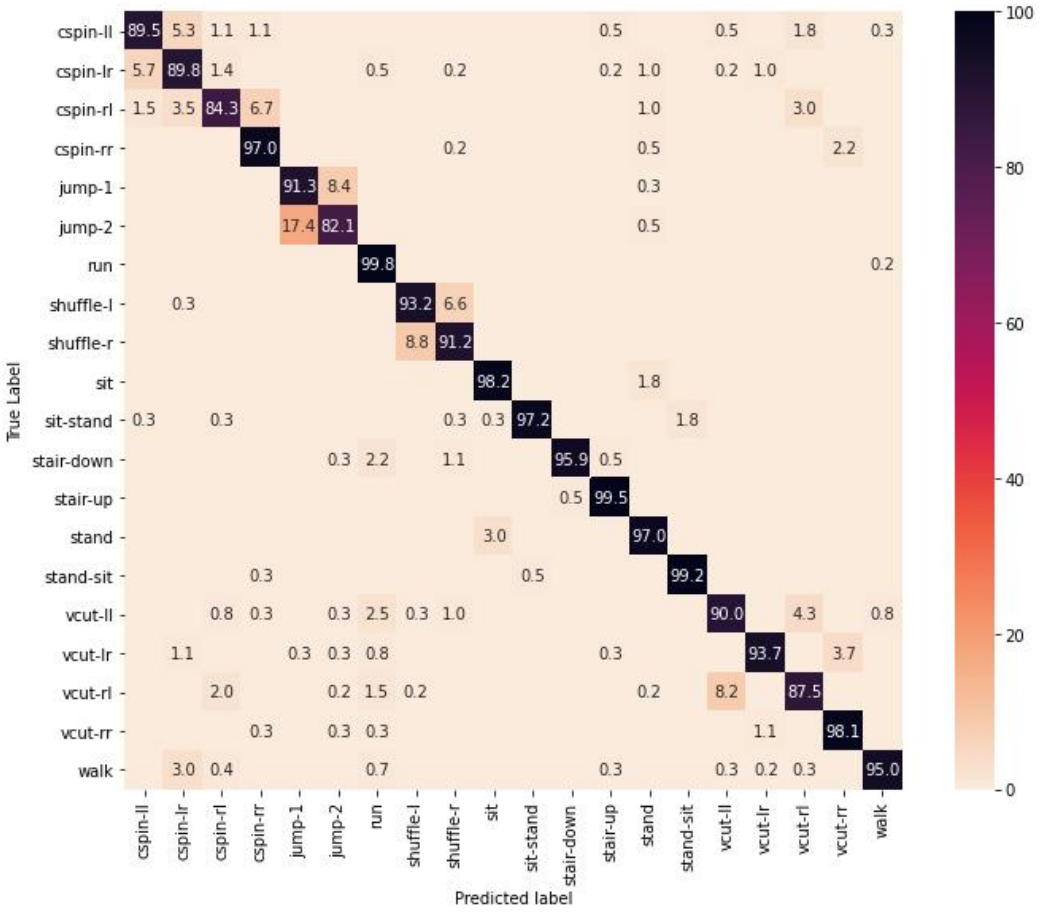


Figure 5.25 – Confusion matrix of cross-validation person-independent recognition results in percentage on the CSL19 dataset: single-state HMM topology.

The statistics convey that by applying only a single HMM, we can already obtain a relatively good recognition accuracy — 94% and 82% for CSL19 and UMS9, respectively, which may be due to the limited number of activities, and the well-prepared offline segmentation.

The "run"/"running" activities in both datasets achieve the best recognition rate among activities. A reasonable explanation is that due to the translational speed and motion intensity, almost every sensor, such as accelerometers, gyroscopes, EMG sensors, and the electrogoniometer, emits significantly discriminative signals compared to other activity's output. For single-state HMM topology, these distinctive signal data undoubtedly reduce the difficulty of training and recognition.

Table 5.9 – Criteria of cross-validation person-independent recognition results on the CSL19 dataset: single-state HMM topology.

Activity	Precision	Recall	F-Score
spin-left-left-first	0.92	0.89	0.91
spin-left-right-first	0.83	0.90	0.86
spin-right-left-first	0.93	0.84	0.88
spin-right-right-first	0.92	0.97	0.94
jump-one-leg	0.84	0.91	0.87
jump-two-leg	0.89	0.82	0.86
run	0.91	1.00	0.95
shuffle-left	0.91	0.93	0.92
shuffle-right	0.90	0.91	0.91
sit	0.97	0.98	0.97
sit-to-stand	0.99	0.97	0.98
walk-downstairs	0.99	0.96	0.98
walk-upstairs	0.98	0.99	0.99
stand	0.95	0.97	0.96
stand-to-sit	0.98	0.99	0.99
V-cut-left-left-first	0.90	0.90	0.90
V-cut-left-right-first	0.97	0.94	0.95
V-cut-right-left-first	0.90	0.88	0.89
V-cut-right-right-first	0.94	0.98	0.96
walk	1.00	0.95	0.97
Global accuracy: 0.94			

Although there is not much room for accuracy enhancement on the CSL19 dataset, there is still an opportunity for further modeling improvement. For instance, "jumping" in the UMS9 dataset has been well recognized (93% accuracy), while the two jump-activities in the CSL19 dataset, "jump-one-leg" and "jump-two-leg," are often confused in the recognition. Based on the values in the confusion matrix, it can be speculated that if we merge these two activities in the CSL19 dataset into one activity named "jump," it should also be recognized as well as in the UMS9 dataset. However, we still want to distinguish them into two activities because they produce different knee loads, which is important for future research of assisting the early treatment of gonarthrosis (see Section 1.2). The current results tell clear that using the single-state HMM topology, "jump" has a relatively high degree of recognition accuracy, but more detailed jump motions are difficult to interpret accurately by a single-state.

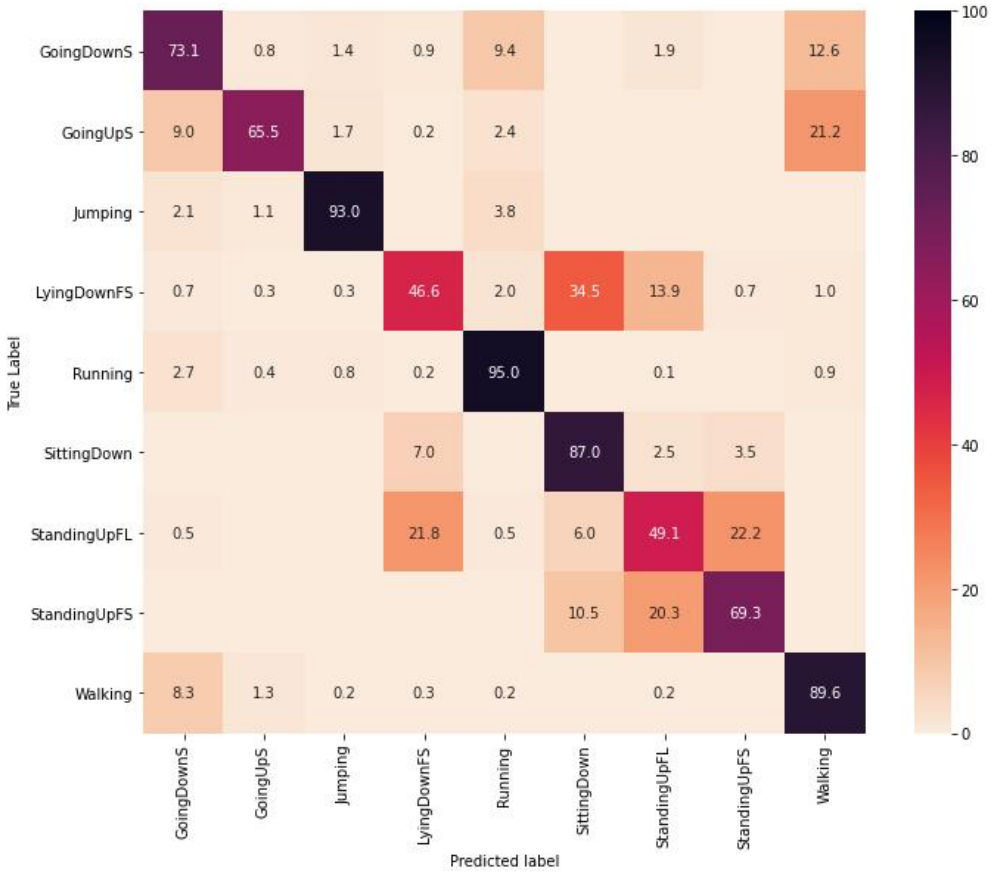


Figure 5.26 – Confusion matrix of cross-validation person-independent recognition results in percentage on the UMS9 dataset: single-state HMM topology.

Table 5.10 – Criteria of cross-validation person-independent recognition results on the UMS9 dataset: single-state HMM topology.

Activity	**Precision**	**Recall**	**F-Score**
GoingDownStairs	0.76	0.73	0.75
GoingUpStairs	0.92	0.65	0.77
Jumping	0.93	0.93	0.93
LyingDownFromStanding	0.62	0.47	0.53
Running	0.91	0.95	0.93
SittingDown	0.57	0.87	0.69
StandingUpFromLaying	0.50	0.49	0.50
StandingUpFromSitting	0.65	0.69	0.67
Walking	0.80	0.90	0.85
Global accuracy: 0.82			

5.6.2 Fixed-Number-of-State Topology

The upper reference of the partitioning and MU experimental study is derived using a six-state topology, which has been proven competent (Hartmann et al., 2020), as introduced in Sections 4.1.1 and 4.1.2.

Figure 5.27 and Table 5.11 describe the confusion matrices and the criteria precision, recall, F-scores, and recognition accuracy of the fixed-number-of-state (six) HMM topology recognition results on the CSL19 dataset, respectively, while Figure 5.28 and Table 5.12 exhibit the results on the UMS9 dataset.

The statistics evidence that compared with single-state, the use of six-state HMM topology in recognition experiments further improves the global recognition accuracy. On the CSL19 and the UMS9 datasets, the accuracy increases by 3% (from 94% to 97%) and 7% (from 82% to 89%), respectively, a considerable enhancement for an already higher single-state benchmark.

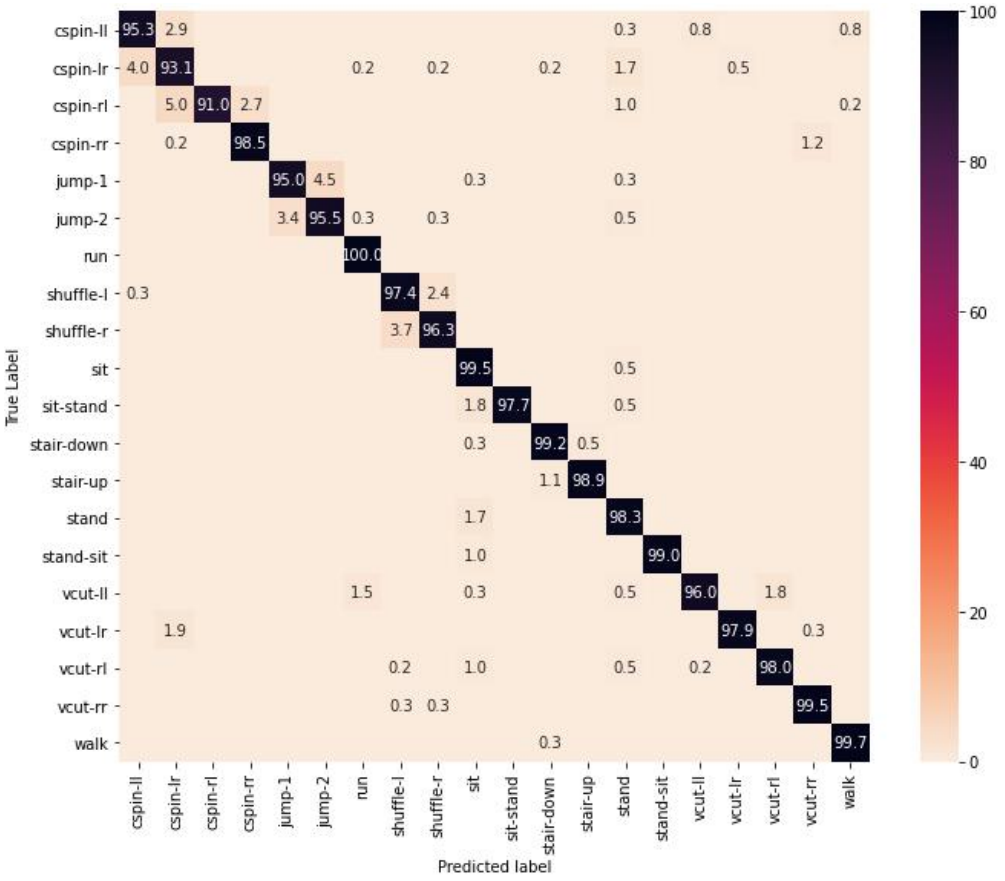


Figure 5.27 – Confusion matrix of cross-validation person-independent recognition results in percentage on the CSL19 dataset: six-state HMM topology.

Table 5.11 – Criteria of cross-validation person-independent recognition results on the CSL19 dataset: six-state HMM topology.

Activity	Precision	Recall	F-Score
spin-left-left-first	0.95	0.95	0.95
spin-left-right-first	0.91	0.93	0.92
spin-right-left-first	1.00	0.91	0.95
spin-right-right-first	0.97	0.98	0.98
jump-one-leg	0.97	0.95	0.96
jump-two-leg	0.96	0.96	0.96
run	0.98	1.00	0.99
shuffle-left	0.96	0.97	0.97
shuffle-right	0.97	0.96	0.97
sit	0.94	0.99	0.97
sit-to-stand	1.00	0.98	0.99
walk-downstairs	0.98	0.99	0.98
walk-upstairs	0.99	0.99	0.99
stand	0.95	0.98	0.96
stand-to-sit	1.00	0.99	0.99
V-cut-left-left-first	0.99	0.96	0.97
V-cut-left-right-first	0.99	0.98	0.99
V-cut-right-left-first	0.98	0.98	0.98
V-cut-right-right-first	0.98	0.99	0.99
walk	1.00	1.00	1.00
Global accuracy: 0.97			

Although for the CSL19 dataset, the recognition accuracy improvement is only 3% (the base is already high), all activities' recognition accuracy in the CSL19 dataset exceeds 90%, which is a qualitative improvement compared to the single-state results. The "run"/"running" activities still lead the way in both datasets, and their recognition rates have risen further. On the CSL19 dataset, the recognition accuracy of "run" even reaches 100%, which means fully correct recognition. The two "jump" activities in the CSL19 dataset are significantly better distinguished from each other, revealing that relatively complex activities like "jump" require more states to achieve more accurate modeling. It is worth mentioning that the recognition accuracy of the two activities in the UMS9 dataset, "LyingDownFromStanding" and "StandingUpFromLaying," both increased from about 40% to more than 70%. Obviously, the use of single-state modeling for these two complex activities involving a series of body movements is imperfect.

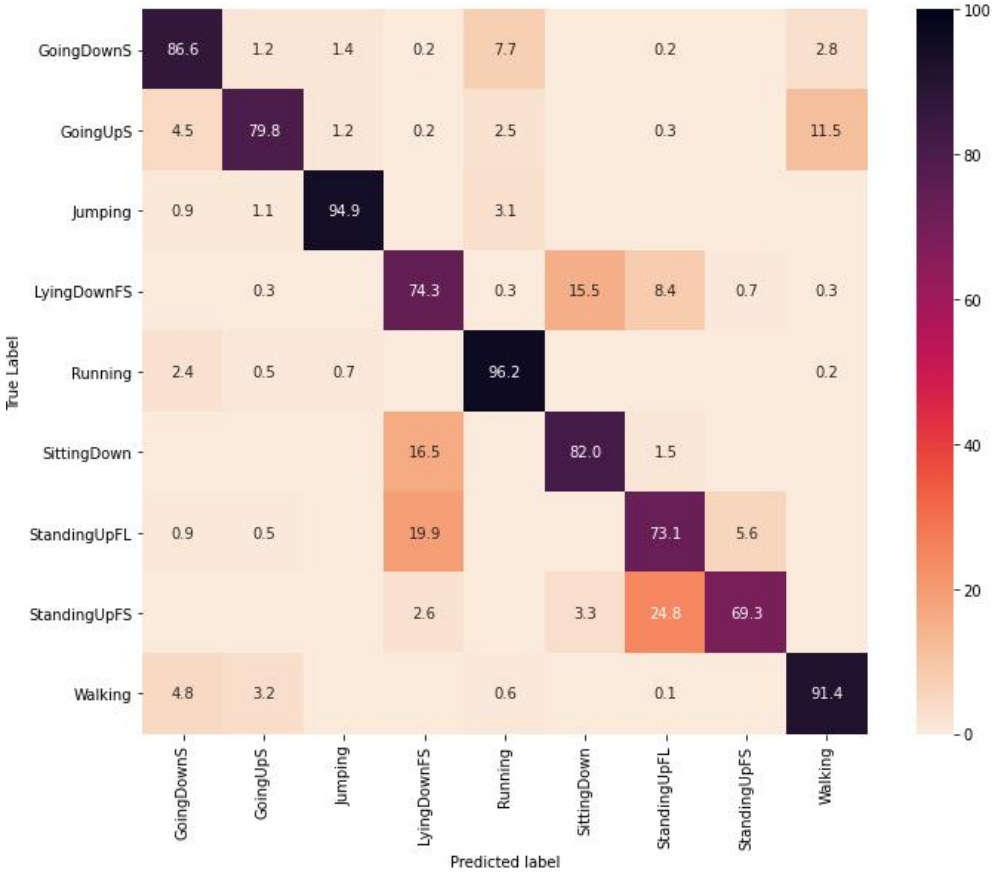


Figure 5.28 – Confusion matrix of cross-validation person-independent recognition results in percentage on the UMS9 dataset: six-state HMM topology.

Table 5.12 – Criteria of cross-validation person-independent recognition results on the UMS9 dataset: six-state HMM topology.

Activity	Precision	Recall	F-Score
GoingDownStairs	0.86	0.87	0.86
GoingUpStairs	0.89	0.80	0.84
Jumping	0.94	0.95	0.95
LyingDownFromStanding	0.72	0.74	0.73
Running	0.92	0.96	0.94
SittingDown	0.76	0.82	0.79
StandingUpFromLaying	0.68	0.73	0.71
StandingUpFromSitting	0.88	0.69	0.78
Walking	0.91	0.91	0.91
Global accuracy: 0.89			

Generally speaking, almost all activities in these two datasets are better recognizable by increasing the number of states from one to six. In the UMS9 dataset, only the recognition accuracy of the activity "StandingUpFromSitting" (i.e., "sit-to-stand") remains almost unchanged, but the criteria precision, recall, and F-Score are improved. In the CSL19 dataset, only "stand-to-sit" undergoes a slightly worse recognition accuracy. Both phenomena mentioned above should be due to the excessive number of HMM states to model the simple body translation, and it can also be described as "unreasonable over-engineering." The emergence of this problem also helps emphasize the importance of the subsequent phase and state partitioning experimental research, trying to model different activities with a different number of states.

5.6.3 Phase and State Partitioning Topology

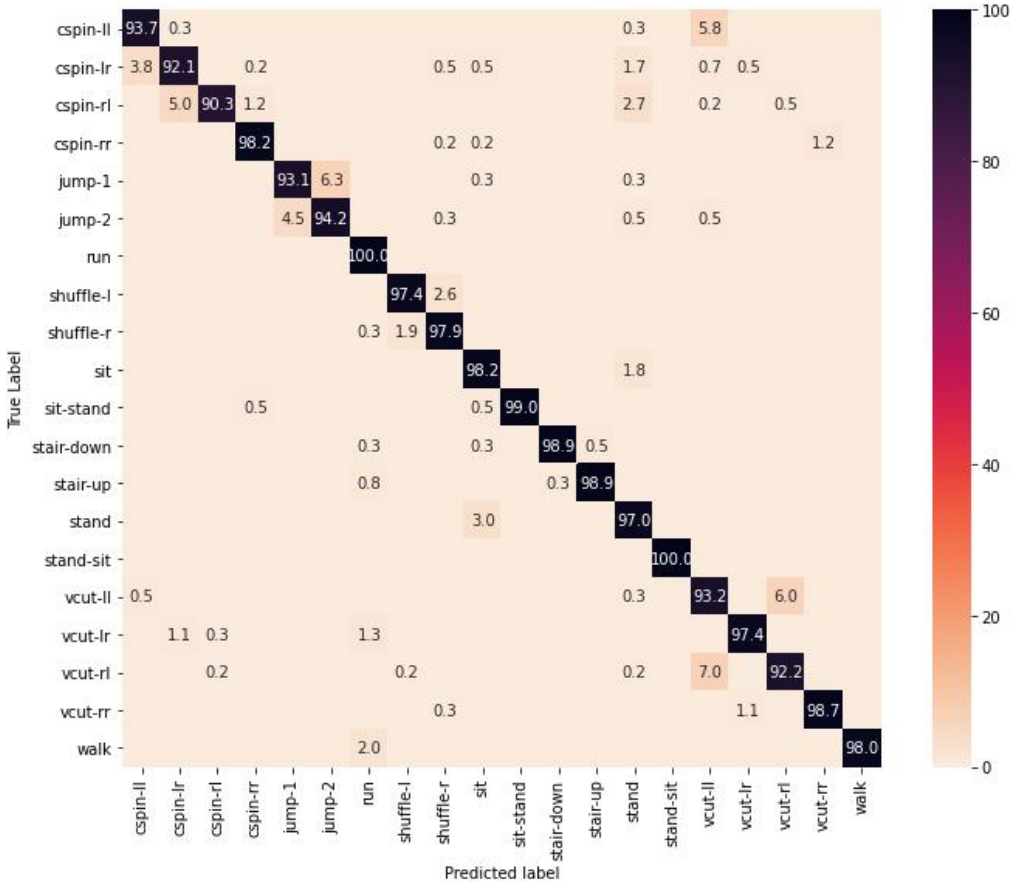


Figure 5.29 – Confusion matrix of cross-validation person-independent recognition results in percentage on the CSL19 dataset: phase and state partitioning HMM topology.

The advanced topology introduces phase and state partitioning, which is described in Section 5.3.

Figure 5.29 and Table 5.13 demonstrate the confusion matrices and the criteria precision, recall, F-scores, and recognition accuracy of the phase and state partitioning HMM topology recognition results on the CSL19 dataset, while Figure 5.30 and Table 5.14 display the results on the UMS9 dataset.

Table 5.13 – Criteria of cross-validation person-independent recognition results on the CSL19 dataset: phase and state partitioning HMM topology.

Activity	Precision	Recall	F-Score
spin-left-left-first	0.96	0.95	0.95
spin-left-right-first	0.94	0.94	0.94
spin-right-left-first	0.98	0.91	0.95
spin-right-right-first	0.99	0.99	0.99
jump-one-leg	0.95	0.93	0.94
jump-two-leg	0.94	0.95	0.94
run	0.96	1.00	0.98
shuffle-left	0.98	0.98	0.98
shuffle-right	0.97	0.98	0.97
sit	0.95	0.98	0.97
sit-to-stand	0.99	0.97	0.98
walk-downstairs	0.99	0.99	0.99
walk-upstairs	0.99	1.00	1.00
stand	0.93	0.97	0.95
stand-to-sit	0.98	0.99	0.99
V-cut-left-left-first	0.96	0.96	0.96
V-cut-left-right-first	1.00	0.99	1.00
V-cut-right-left-first	0.97	0.96	0.97
V-cut-right-right-first	0.99	1.00	0.99
walk	1.00	0.99	0.99
Global accuracy: 0.97			

Our experiments show that phase and state partitioning can retain (97% on the CSL19 dataset) and even improve (90% on the UMS9 dataset) recognition accuracy and class separability.

The change in the number of states will also cause a change in the number of trainable units and Gaussians, which will be analyzed in Section 5.6.4.

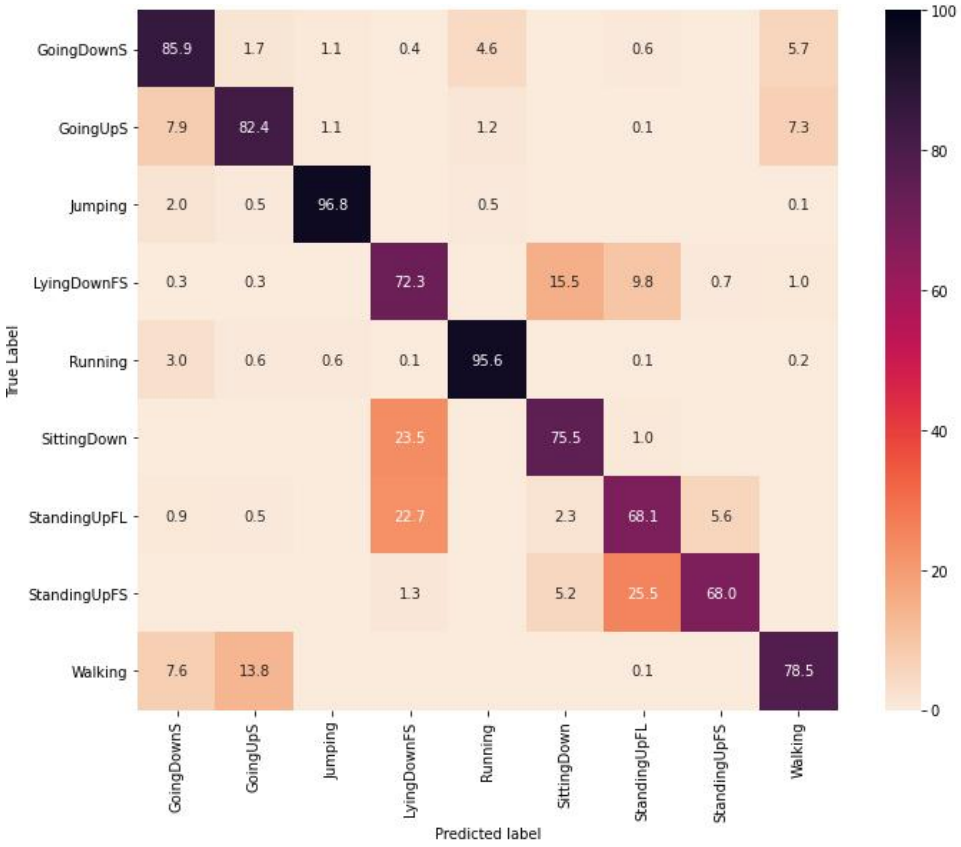


Figure 5.30 – Confusion matrix of cross-validation person-independent recognition results in percentage on the UMS9 dataset: phase and state partitioning HMM topology.

Table 5.14 – Criteria of cross-validation person-independent recognition results on the UMS9 dataset: phase and state partitioning HMM topology.

Activity	**Precision**	**Recall**	**F-Score**
GoingDownStairs	0.83	0.91	0.87
GoingUpStairs	0.89	0.88	0.88
Jumping	0.96	0.97	0.96
LyingDownFromStanding	0.71	0.71	0.71
Running	0.96	0.96	0.96
SittingDown	0.76	0.84	0.80
StandingUpFromLaying	0.63	0.71	0.67
StandingUpFromSitting	0.89	0.66	0.76
Walking	0.95	0.89	0.92
Global accuracy: 0.90			

In the experiments on the CSL19 dataset, "run" still maintains a 100% recognition accuracy, and "stand-to-sit" has become the second completely accurately recognized activity. In Section 5.6.2, we found that the recognition rate of "stand-to-sit" decreased after the number of states changed from one to six. However, if we take the dual-state design based on the phase and state partitioning experimental study, it can be "perfectly" recognized, highlighting the importance and effectiveness of the phase and state partitioning study.

The experimental results of phase and state partitioning are encouraging and will serve as a baseline for MU-based generalization experiments. As mentioned in Section 5.5.2, the robustness of the current partitioning baseline helps carry out the MU-based generalization attempts, experiments, and design iterations more efficiently with less effort.

5.6.4 MU-Based Generalized Topology and Joint Analysis between Topologies

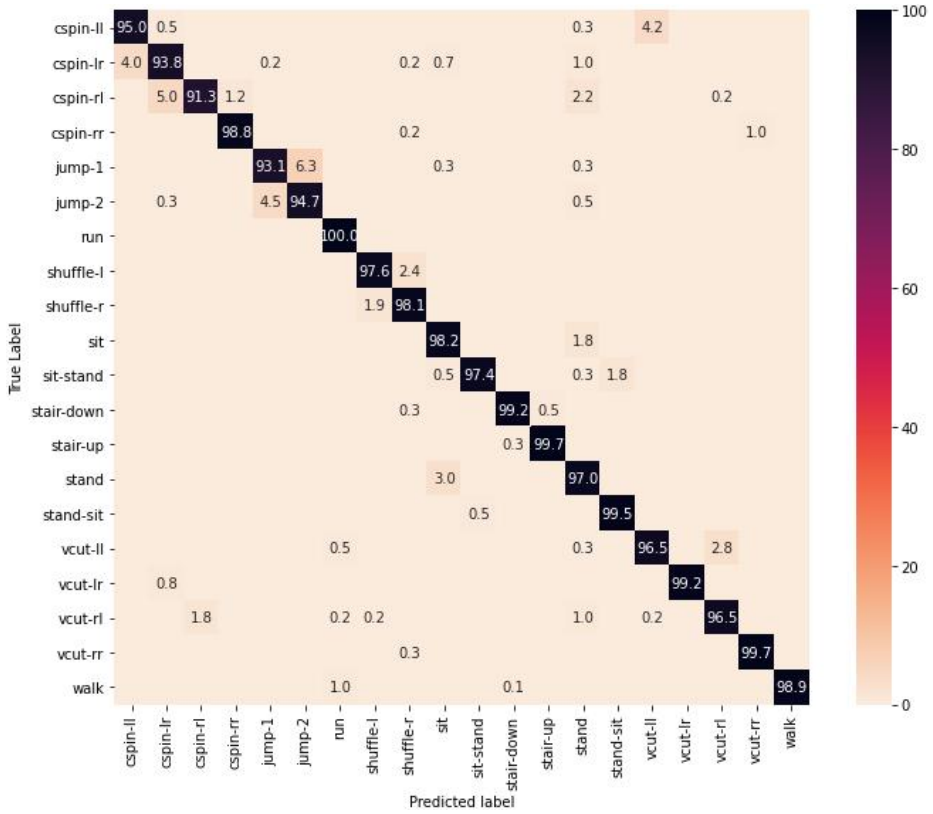


Figure 5.31 – Confusion matrix of cross-validation person-independent recognition results in percentage on the CSL19 dataset: MU-based HMM topology.

The topology with MU-based generalization shares states between activities and implements MUs based on the generalization study approaches described in Section 5.5.2. The detailed generalization scheme can be found in Section 5.7.3.

Figure 5.31 and Table 5.15 give the confusion matrices and the criteria precision, recall, F-scores, and recognition accuracy of the MU-based generalized HMM topology recognition results on the CSL19 dataset, respectively, while Figure 5.32 and Table 5.16 state the results on the UMS9 dataset.

Table 5.15 – Criteria of cross-validation person-independent recognition results on the CSL19 dataset: MU-based HMM topology.

Activity	**Precision**	**Recall**	**F-Score**
spin-left-left-first	0.95	0.94	0.94
spin-left-right-first	0.94	0.92	0.93
spin-right-left-first	0.99	0.90	0.95
spin-right-right-first	0.98	0.98	0.98
jump-one-leg	0.95	0.93	0.94
jump-two-leg	0.94	0.94	0.94
run	0.92	1.00	0.96
shuffle-left	0.98	0.97	0.98
shuffle-right	0.96	0.98	0.97
sit	0.95	0.98	0.97
sit-to-stand	1.00	0.99	0.99
walk-downstairs	1.00	0.99	0.99
walk-upstairs	0.99	0.99	0.99
stand	0.93	0.97	0.95
stand-to-sit	1.00	1.00	1.00
V-cut-left-left-first	0.87	0.93	0.90
V-cut-left-right-first	0.98	0.97	0.98
V-cut-right-left-first	0.93	0.92	0.93
V-cut-right-right-first	0.99	0.99	0.99
walk	1.00	0.98	0.99
Global accuracy: 0.97			

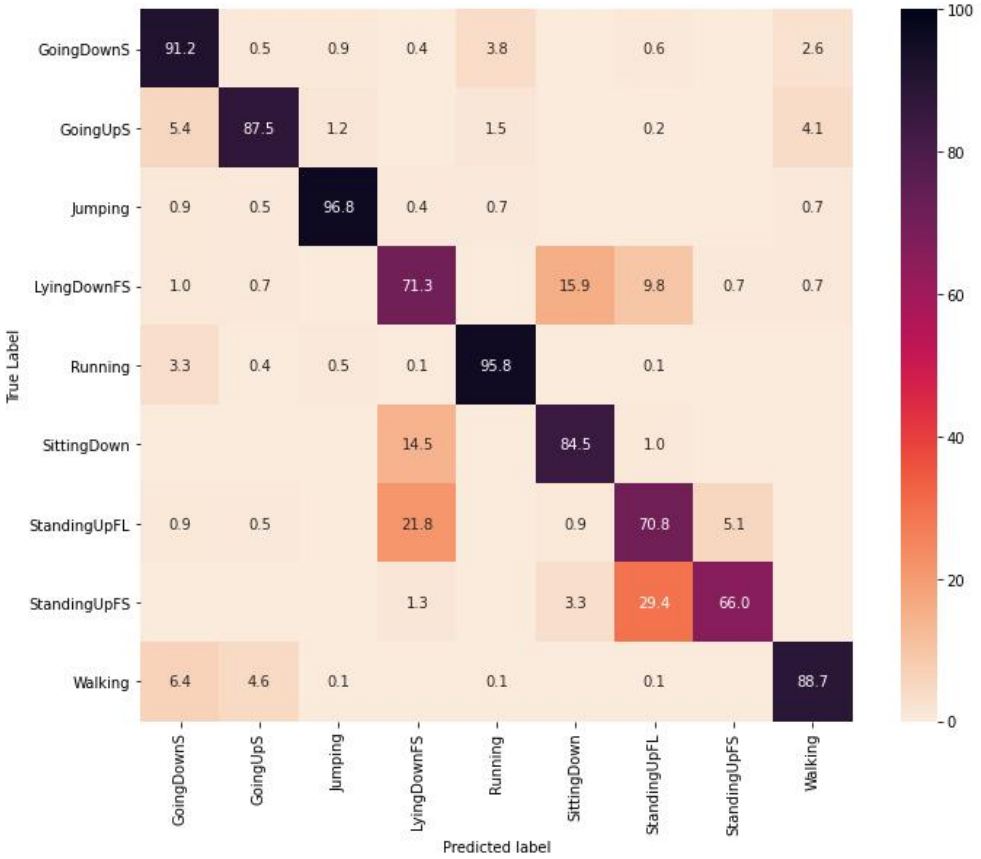


Figure 5.32 – Confusion matrix of cross-validation person-independent recognition results in percentage on the UMS9 dataset: MU-based HMM topology.

Table 5.16 – Criteria of cross-validation person-independent recognition results on the UMS9 dataset: MU-based HMM topology.

Activity	**Precision**	**Recall**	**F-Score**
GoingDownStairs	0.80	0.86	0.83
GoingUpStairs	0.73	0.82	0.77
Jumping	0.95	0.97	0.96
LyingDownFromStanding	0.67	0.72	0.70
Running	0.96	0.96	0.96
SittingDown	0.72	0.76	0.74
StandingUpFromLaying	0.64	0.68	0.66
StandingUpFromSitting	0.88	0.68	0.77
Walking	0.90	0.79	0.84
Global accuracy: 0.86			

Additional to the above-described HMM-based recognition experiments of different topologies in Section 5.6, an independent evaluation is performed on the forced aligned and re-labeled data for each a *k-Nearest Neighbors* (kNN) algorithm , LDA, and *Nearest Centroid* (NC) classifier. Table 5.17 summarizes all the experimental results of all four types of HMM topology and the LDA evaluation. The kNN and NC classifiers are omitted for brevity, as they highly correlate with the LDA evaluation.

Table 5.17 – Summary of MU-based experimental results on the CSL19 and the UMS9 datasets. #: number of.

Dataset	Model	#Units	#Gaussians	HMM	LDA
CSL19	Fixed	120	840	97.5%	23.3%
	Partition	84	588	97.3%	29.3%
	MU	60	420	96.5%	32.9%
	Single	20	140	93.6%	24.1%
UMS9	Fixed	54	365	88.9%	16.3%
	Partition	60	378	89.9%	26.4%
	MU	41	281	85.7%	26.1%
	Single	9	63	82.2%	26.2%

Table 5.17 reads as follows: the number of states denotes the sum of individual states across all activities and denotes the number of target classes for the LDA to learn. The number of Gaussians denotes the sum of unique Gaussians and, in combination with the number of units, indicates the number of trainable parameters. The HMM column denotes recognition accuracy over activities, while the LDA accuracy provides an indirect measure of state separability and is not directly comparable to the HMM accuracy.

From Table 5.17 we can conclude that phase and state partitioning can retain and even improve recognition accuracy and class separability with fewer (30% less on the CSL19 dataset) or similar (13 more on the UMS9 dataset) parameters as the commonly applied fixed-number-of-state topology. Note that after applying phase and state partitioning to the UMS9 dataset, the total number of units and Gaussians has increased compared to the six-state topology; however, from Table 5.6, we can find that the number of states we apply to the UMS9 dataset should be less than the fixed-number-of-state (six) topology ($6 \times 9 = 54$). The reason is that in the UMS dataset, the precise number of gait cycles per segment varies due to the constant segment length of 3 seconds. Therefore, a single gait cycle is modeled, and an initial and terminal "random"-state which consumes all other data added.

Our MU-based generalized HMM topology further reduces the number of parameters by 30% over the partitioned model while slightly decreasing recognition accuracy (0.8% on the CSL19 dataset and 4.2% on the UMS9 dataset). If the maximum number of Gaussians in the *Split-and-Merge* training is increased from 7 to 20, the accuracy only decreases by 0.7% (on the CSL19 dataset) and 1.4% (on the UMS9 dataset). In continuation of the partitioning experiments, state separability is increased further. At the expense of a bit of recognition accuracy, the trade-off is a significant simplification of activity modeling as well as state interpretability.

Too few trainable parameters might explain the performance drops. However, a further model iteration is required to prohibit shared states from capturing several distinct phases simultaneously by utilizing the additional parameters.

5.7 MU-DNA (Directional Nomenclature + Anchored) and MU-Gene (Generalization)

As the results in Section 5.6.4 confirms, MU opens up a new and effective way for human activity modeling. Therefore, MUs can play an essential role in efficiently modeling new activities and should be given interpretable and meaningful names.

5.7.1 Methodology

We do not recommend using numbers as a distinction between two different MUs. Numbers should describe physical quantities, such as speed levels and angles. To define an MU with a specific motion trial, we propose a *Six-Directional Nomenclature* (6DN). In brief, 6DN means the six directions front (forwards), back(wards), up(wards), down(wards), left(wards), and right(wards) in the torso coordinate system, and their various combinations. We use the letters **F**, **B**, **U**, **D**, **L**, and **R** to abbreviate them, respectively.

The theoretical and experimental research in the example of "walk" versus "walk-upstairs," as detailed in Section 5.5.2, results in them sharing the first two MUs, **walk-ISt-F** and **walk-MSt-F**. The remaining third, fourth, and fifth MUs are defined as follows, respectively: "walk": **walk-TSt-F**, **walk-ISw-F**, and **walk-TSw-F**; "walk-upstairs": **walk-TSt-FU**, **walk-ISw-FU**, and **walk-TSw-FU**. It is noticeable that "walk" in the above-mentioned MUs represents the primary category of these two activities (they are both

gait-based activities; "walk-upstairs" is a walk-based activity with a unique direction), **St** and **Sw** indicate the phases, and **I**, **M**, and **T** denote the sub-phases, as described in Section 5.3.4. **F** (front) and **FU** (front+up) are related to more exclusive states within 6DN to distinguish different MUs according to movement directions.

All activities described by 6DN can be decomposed into one or several axial translational movements, which will cause the body or body part to leave its original position. If an activity or an MU does not involve (or if we define them without the intention for) translational movement, we can regard them as "an **Anchored** activity/MU" and do not need to use 6DN to define them. An obvious example is the activity "stand" and its single MU **stand**, to which adding any direction is superfluous.

In summary, for any activity and its attached MUs, they either have translational movement (defined by 6DN) or not (Anchored). We abbreviate such a "Motion Units' Directional Nomenclature/Anchored" pattern as the "**MU-DNA**" of human activities. Moreover, considering the literal aesthetics, we can also abbreviate the "Motion Units' Generalization" as "**MU-Gene**," making it picturesque to think of MUs as MU-Gene and MU-DNA (different from the biological meaning).

5.7.2 Modeling Human Activities Using MU-Gene and MU-DNA

The **Anchored** activities in the activity's MU-DNA pattern are usually modeled with one discriminating MU. Moreover, we can quickly model new activities preliminarily based on the 6DN in the activity's MU-DNA pattern. For example, "V-cut-left" in the CSL19 dataset is also a gait-based activity that can be analogically modeled as {**run-ISt-F**, **run-MSt-F**, **run-TSt-FL**, **run-ISw-FL**, **run-TSw-FL**}, and "jump" (upwards) in both CSL19 and UMS datasets is modeled as {**jump-ITo-U**, **jump-TTo-U**, **jump-Sh-U**, **jump-ILa-D**, **jump-TLa-D**}. A uniquely defined activity "LyingDownFromStanding" in the UMS dataset is partitioned into five MUs {**ISh-D**, **MSh-D**, **TSh-D**, **lie-I**, **lie-T**}.

Table 5.18 lists some categories of ambulation activities and the suggested number of MUs to model them as a universal template. After investigating the number of states for each activity, i.e., phase and state partitioning (see Section 5.3), we can refer to the merging strategy and repeated experiments described in Section 5.5.2 to study the MU-Gene of all states and define them

Table 5.18 – Some categories of a universal activity modeling template of ambulation activities and suggested number of MUs for modeling them.

Activity Category	Activity examples	#MUs
Anchored	sit, stand, lie, squat, tiptoe	1
Gait-based	walk, march on the spot	5 per gait
Jump-based	jump, flip, dive	5 or 5+
In-place simple translation	sit-to-stand, stand-to-sit	2 or 3
In-place complex translation	lay-to-stand, stand-to-lay	5 or 5+

using MU-DNA. Thus, we will finally have the MU design of the complete set of activities in each dataset.

For another example of expandability, an artificially defined football-specific training activity "save a penalty by jumping right-forward" may be modeled by combining "jump" and the directions **FUR** and **FDR** of 6DN in the activity's MU-DNA pattern, as illustrated in Figure 5.33.

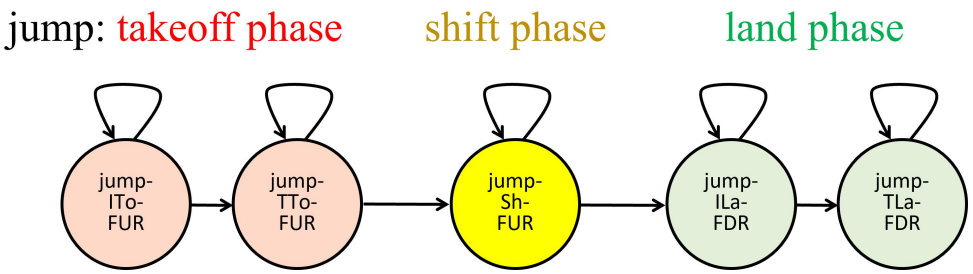


Figure 5.33 – A linear left-right HMM for the three-phase-five-state artificially defined football-specific activity "save a penalty by jumping right-forward" based on MU-DNA. Red: states/sub-phases in the takeoff phase; yellow: state/sub-phase in the shift phase; green: states/sub-phases in the land phase.

Rapid and straightforward modeling does not make up the entirety of model research. Adequate and appropriate repeated verification experiments and parameter tuning, such as described in many of the above sections, are still essential.

5.7.3 Human Activity Modeling Design on the CSL19 and the UMS9 Datasets Based on MU-Gene and MU-DNA

Figures 5.34—5.36 illustrate MU-based generalized linear left-right HMMs for each activity in the CSL19 and the UMS9 datasets of the optimized experimental performance so far (see Section 5.6.4) by utilizing MU-Gene and MU-DNA.

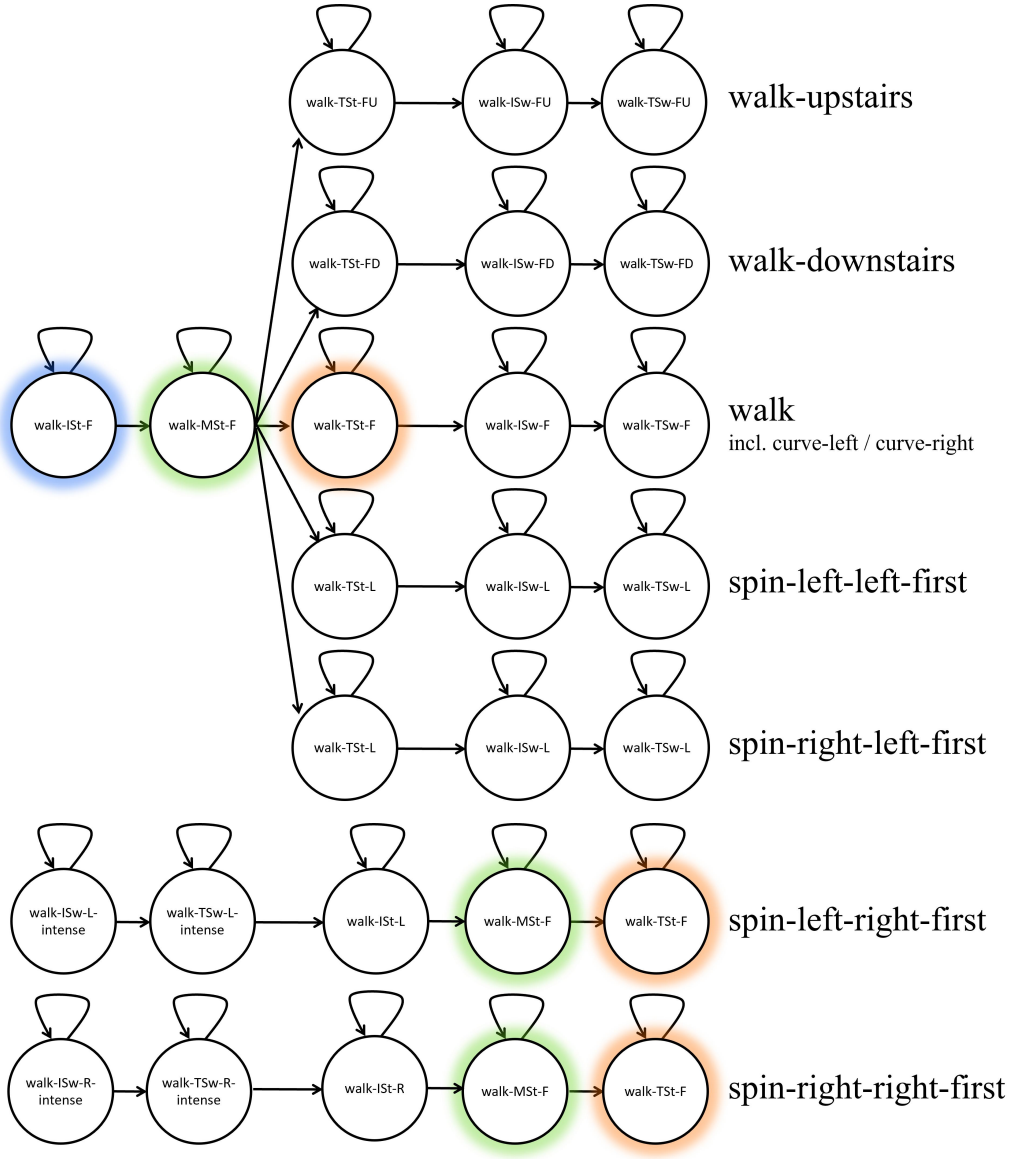


Figure 5.34 – The MU-based generalized linear left-right HMMs for each light gait-based activity in the CSL19 and the UMS datasets. Background colors of certain states: MUs with reusability in the datasets.

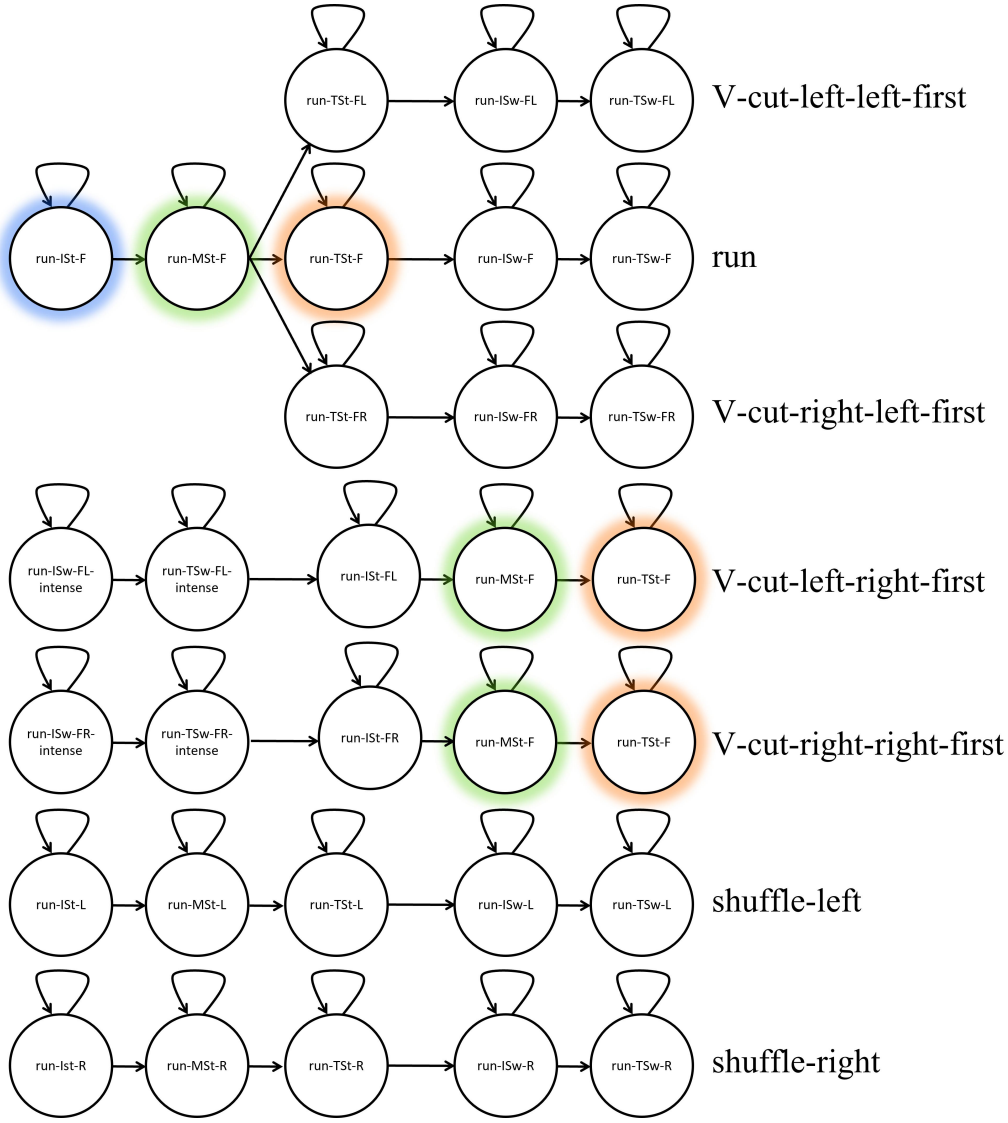


Figure 5.35 – The MU-based generalized linear left-right HMMs for each intensive gait-based activity in the CSL and the UMS datasets. Background colors of certain states: MUs with reusability in the datasets.

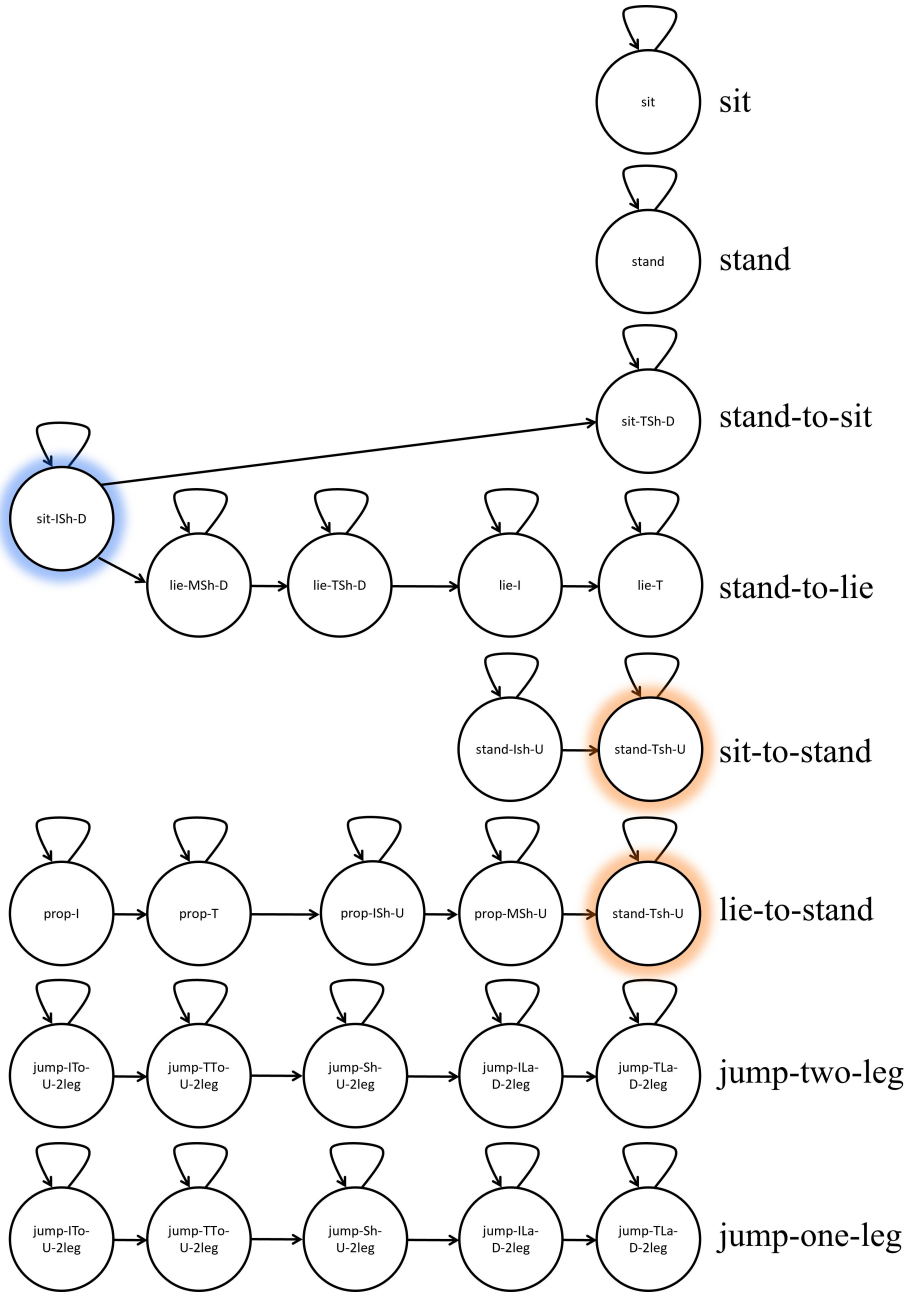


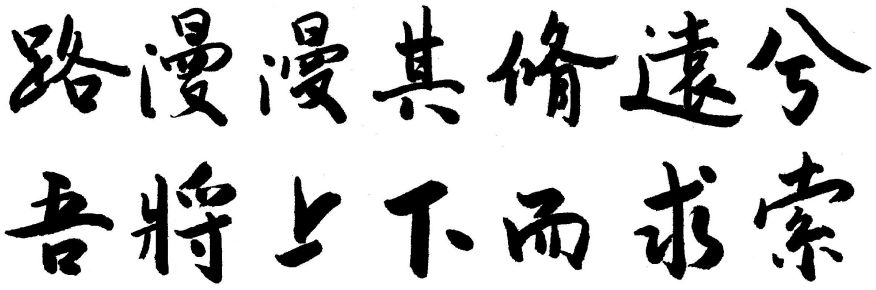
Figure 5.36 – The MU-based generalized linear left-right HMMs for jumping, sitting, standing, and lying related activities in the CSL and the UMS datasets. Background colors of certain states: MUs with reusability in the datasets.

According to Table 5.17, the MU design with MU-Gene and MU-DNA reduced CSL19's number of HMM states from 84 (phase and state partitioning) to 60, a 28.6% reduction. For UMS9, the number of states dropped from 60 to 41, a reduction of 35%.

Noticeably, the same activity in both CSL19 and UMS datasets are modeled by the same MUs, despite that the two datasets apply different sensors, sensor positioning, activity definitions, and segmentation methods. The experimental results (see Section 5.6) show that the MU-based generalization is reasonable, feasible, and extendable.

Chapter 6

Conclusions and Future Work



"*Long, long had been my road and far, far was the journey;*

I would go up and down to seek my heart's desire."

Qu Yuan (c. 340 BC — 278 BC), *Encountering Sorrow (Li Sao)*,

translated by David Hawkes (Hawkes and Qu, 1985).

6.1 Summary of Results

This dissertation's primary goal was to systematically study human activity recognition and enhance its performance by advancing human activities' sequential modeling based on HMM-based machine learning. Driven by these purposes, this dissertation has the following major contributions:

- The proposal of our HAR research pipeline that guides the building of a robust wearable end-to-end HAR system and the implementation of the recording and recognition software ASK according to the pipeline;
- Collecting several datasets of multimodal biosignals from over 25 subjects using the self-implemented ASK software and implementing an easy mechanism to segment and annotate the data;
- The comprehensive research on the offline HAR system based on the recorded datasets and the implementation of an end-to-end real-time HAR system;
- A novel activity modeling method for HAR, which partitions the human activity into a sequence of shared, meaningful, and activity distinguishing states, called *Motion Units* (MUs), analog to phonemes in speech recognition.

6.1.1 Research Pipeline, Software and Datasets

A comprehensive pipeline of nine blocks has been proposed to describe our HAR research as a whole and point out the order and relationship between each block.

Based on the pipeline, the baseline multifunctional ASK software for the HAR study and the series of upgraded and expanded versions were implemented. Their practicability and robustness have been repeatedly verified in various experiments.

We collected several data corpora of multimodal biosignals by applying the functionalities of data acquisition and the protocol for segmentation and annotation in the ASK software. The latest CSL19 dataset is currently being publicized to share the data with researchers in the same or similar fields, based on the fact that its applicability and comprehensiveness has been proven by post verification and various application cases.

6.1.2 HAR Research and MU-Based Human Activity Modeling

HAR research involves various aspects, such as parameter tuning and feature dimensionality reduction. The experimental results based on several datasets have confirmed that our offline recognizer achieves recognition accuracies which are on par with bleeding edge results from the literature. On this basis, our end-to-end real-time HAR system has also achieved good preliminary results and is operating stably.

To better and efficiently model human activities, we propose an activity modeling method based on *Motion Units* (MUs), which has good operability, scalability, and meaningfulness. Using in-house and external datasets to verify the unified MU design, the results indicate that MU design reduces the redundancy of the entire modeling architecture as a whole and improves class separability, laying the foundation for efficient training. Most importantly, MUs are well-defined, meaning-carrying, and activity-composing recognizable units for HAR.

6.2 Perspectives and Future Directions

As one of this dissertation's most essential research objects, machine learning methods have been applied to study the MU design, which opens several new aspects. The easy extension allows expansion to further sensor setups, new activities, and additional datasets.

6.2.1 Potential Future Research on MU Design

Heuristic MU-based modeling template Although MU models of human daily ambulation activities were provided in this dissertation, the MU design aims to cope universally with omnifarious human activities. Just as the saying goes, "*teaching someone how to fish is better than just giving him a fish.*" We plan to build a heuristic modeling template helping researchers more efficiently use MUs to model activities for their datasets, application scenarios, or activity definitions. This architecture can be a block diagram, containing a series of Yes/No questions, guiding the user to determine step by step the number of MUs for each activity, as well as each MU's MU-Gene and MU-DNA.

Body-part activity MUs Among all human activities, the lower limbs' movements, such as walking, running, and jumping, are often more

dominant because they generally determine the entire body's translation and balance. In this regard, this dissertation provides the modeling suggestion of most ambulation activities. The follow-up research can apply MUs to distinguish detailed activities on specific body parts, such as hand/shoulder/head/upper-body activities. For example, the *Upper-body movements* dataset (Santos et al., 2020) contains upper-body-specific activities like flexion/extension/abduction/adduction of the forearm, arm, torso, or wrist, which is applicable as one of the data resources for the body-part MU research.

MU design for continuous activity sequences Many application scenarios or datasets involve continuous activity sequences, like "cooking," "watching television," and "driving home," such as the *PAMAP2* dataset (Reiss and Stricker, 2012b), (Reiss and Stricker, 2012a) and the *Opportunity* dataset (Roggen et al., 2010), (Chavarriaga et al., 2013). On the basis of the low-level activity modeling, whether MU design is adjustable to recognize these higher-level continuous daily activity sequences is an open question.

MUs' application on more public datasets This research is an agglomeration of the above three follow-up research directions. Different datasets will involve different activity definitions, different segmentation methods of time series, different activities on various body parts, and different applications of miscellaneous types of wearable sensors. A broader universal MU architecture is worth further exploring. In addition to the UniMiB SHAR dataset (UMS) used in this dissertation and the three public datasets mentioned above, many multi-sensor, fully-synchronized, and well-segmented datasets, such as *Enabl3s* (Hu et al., 2018), *RealWorld* (Sztyler and Stuckenschmidt, 2016), and *Gait Analysis DataBase* (Loose et al., 2020), can provide more stages for MUs to play and help improve the above-mentioned heuristic MU-based modeling template's applicability.

MU-based real-time HAR system Our current real-time HAR system (ASKED) and the extended plug-and-play add-on (ASKPAPA) are based on the simplest single-state HMM of activity modeling. If we integrate MUs into this real-time recognition system, we can foresee that it will manage to recognize more typical activities and may perform better. How to apply MUs to real-time systems is an extensive research aspect involving accuracy-delay trade-offs, parameter tuning (such as window length and overlap length), and other research topics.

6.2.2 Potential Future Research on Other Aspects of HAR

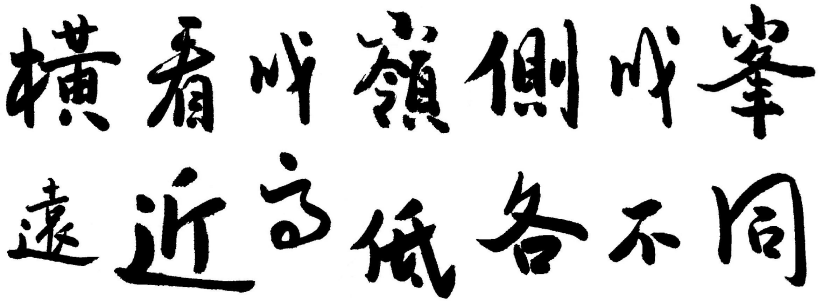
Research on walking with slight directional changes As explained in Section 5.3.5, in our current partitioning and generalization researches, "walk-curve-left (90°)" and "walk-curve-right (90°)" are merged into "walk" because the distinction between these three activities is unsound according to a large number of experimental results. Also, we haven't found any other datasets or literature at present, distinguishing the "in-progress small-angle turn while walking" as an independent activity. Hence, we have merged "walk," "walk-curve-left," and "walk-curve-right" in the latest version. However, the reasonable biosignal-based distinction and robust recognition between these three activities are worth researching.

More potential wearable sensors for HAR Some literature suggests that acoustic sensors may be applicable for estimating knee health. As introduced in Section 2.2.5, many potential sensors, such as acoustic sensors (airborne microphones or piezoelectric sensors), force sensors, barometers, and magnetometers, can be further studied to improve the HAR system.

Combination of video- and biosignal-based HAR As introduced in Section 2.1, video cameras have been widely used in the HAR systems. Some research work, like *EASE* (Meier et al., 2018), (Mason et al., 2018), (Mason et al., 2020), combines video-based and biosignal-based HAR for comparative research and collaborative modeling. This topic of "external + internal sensing for HAR" is both compelling and necessary because it will be widely applicable in intelligent homes and *Fourth Industrial Revolution* (Industry 4.0).

Appendix A

Duration Histograms of All Activities in the CSL19 Dataset



"*From the side, a whole range; from the end, a single peak;*

far, near, high, low, no two parts alike."

Su Shi (1037 — 1101), *Written on the Wall at West Forest Temple*,

translated by Burton Watson (Watson, 1994).

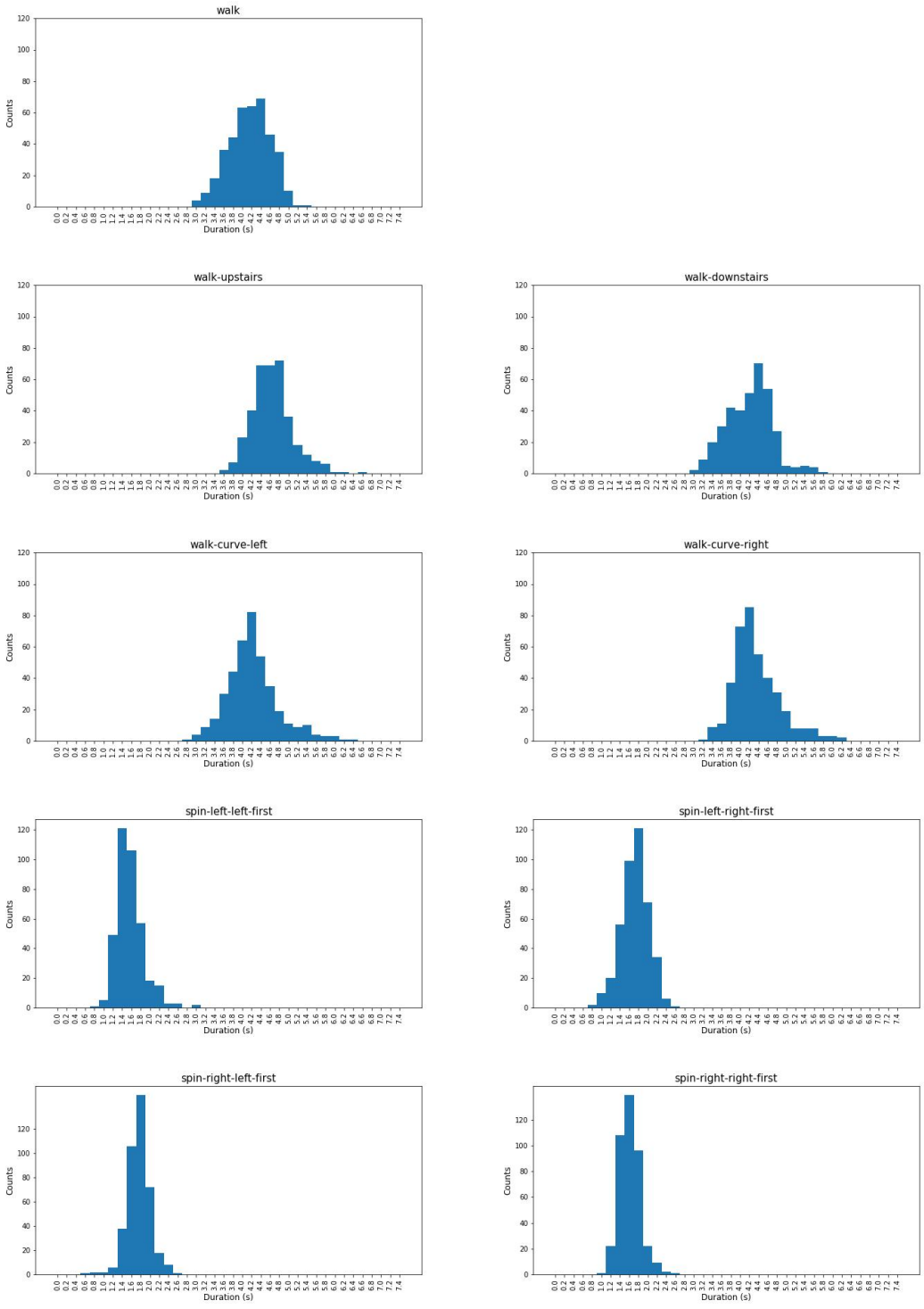


Figure A.1 – Histograms of the activity duration in the CSL19 dataset: gait-based activities. The area under the curve equals the total number of segment occurrences within 200-millisecond intervals.

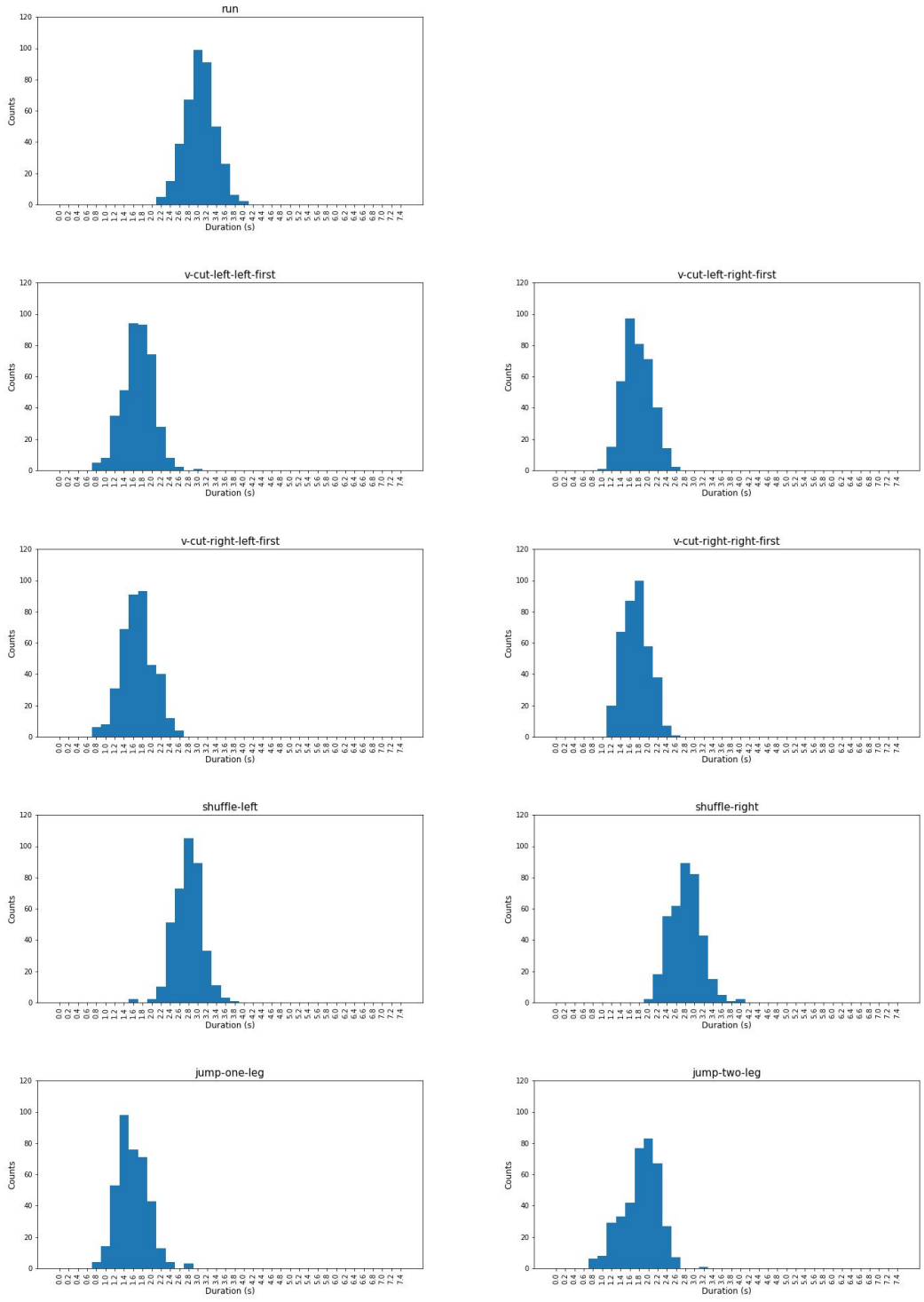


Figure A.2 – Histograms of the activity duration in the CSL19 dataset: intensive gait-based activities and jumps. The area under the curve equals the total number of segment occurrences within 200-millisecond intervals.

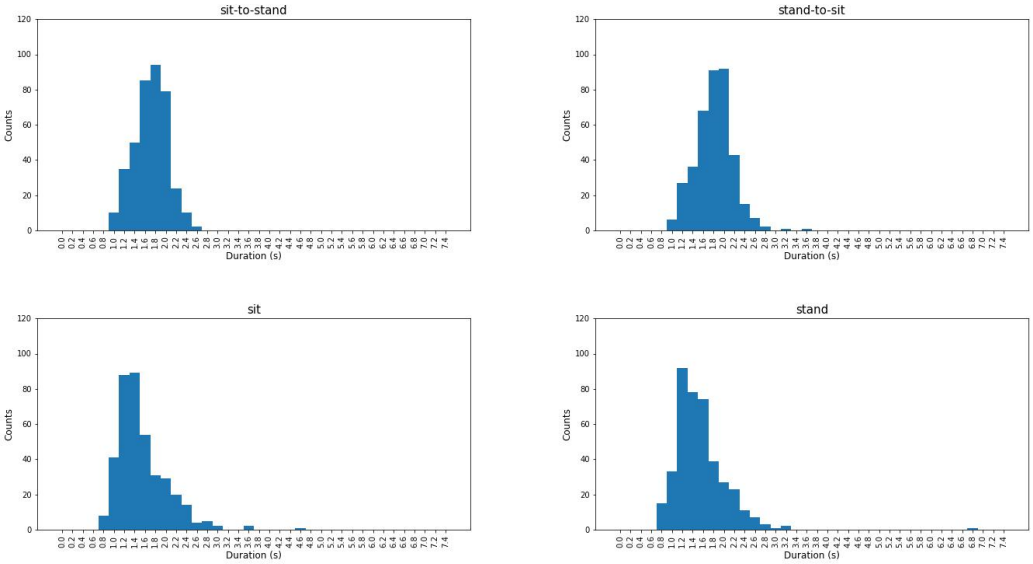


Figure A.3 – Histograms of the activity duration in the CSL19 dataset: "sit," "stand," "sit-to-stand," and "stand-to-sit." The area under the curve equals the total number of segment occurrences within 200-millisecond intervals.

Appendix B

Exemplar Pages of the CSL19 Sensor Data Documentation

"*For scholars, practice is more valuable than theoretical knowledge.*"

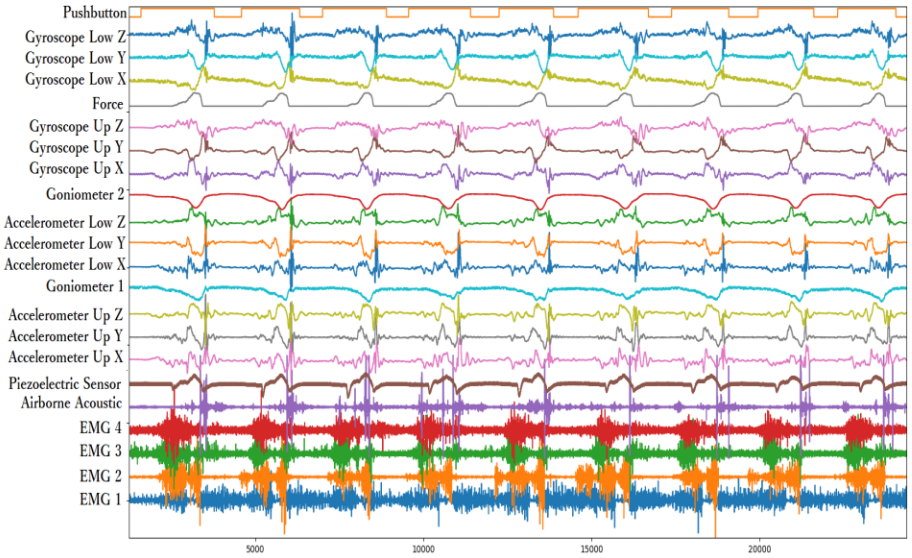
Sima Guang (11/1019 — 10/1086), *Reply to Kong Wenzhong's letter.*

Activity Signal Kits

Sensors

- 4 X EMG
- 4 X IMUs
  - 2 X 3D Accelerometers
  - 2 X 3D Gyroscopes
- 1 X 2D Goniometer
- 1 X Force sensor
- 1 X Piezoelectric sensor
- 1 X Airborne acoustic sensor / microphone
- 1 X Pushbutton for segmentation and annotation

Sensor Legend



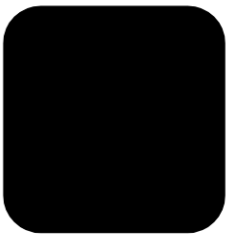
spin-right--left-first: all sensors, nine gait cycles

Created by Hui Liu

Figure B.1 – Exemplar of the sensor data documentation: Page 1.

Profile

File No.:
Name:
Gender:
Age:
Occupation:
Knee Surgery: No

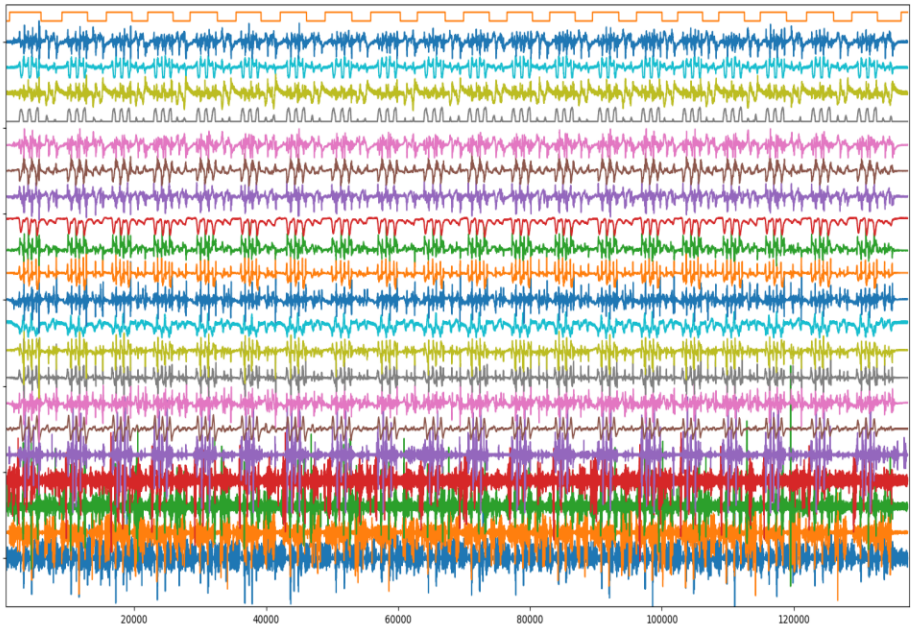


Date:	[redacted]
#Activities Recorded:	22
#Repetition per Activities:	20
#Recording Sets:	17
#Valid Segments with Annotation:	440
Sequence Standardization:	Yes

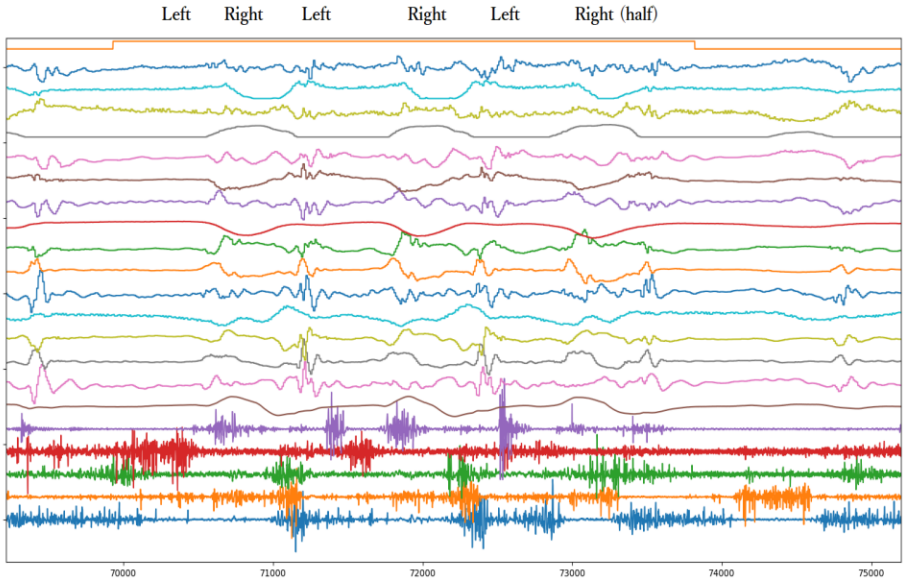
Activities:

- walk
- walk-curve-left, spin-left-left-first, spin-left-right-first
- walk-curve-right, spin-right-left-first, spin-right-right-first
- run
- V-cut-left-left-first, v-cut-left-right-first
- V-cut-right-left-first, v-cut-right-right-first
- shuffle-left, shuffle-right
- sit, sit-to-stand, stand, stand-to-sit
- jump-two-leg, jump-one-leg
- walk-upstairs, walk-downstairs

Figure B.2 – Exemplar of the sensor data documentation: Page 2.

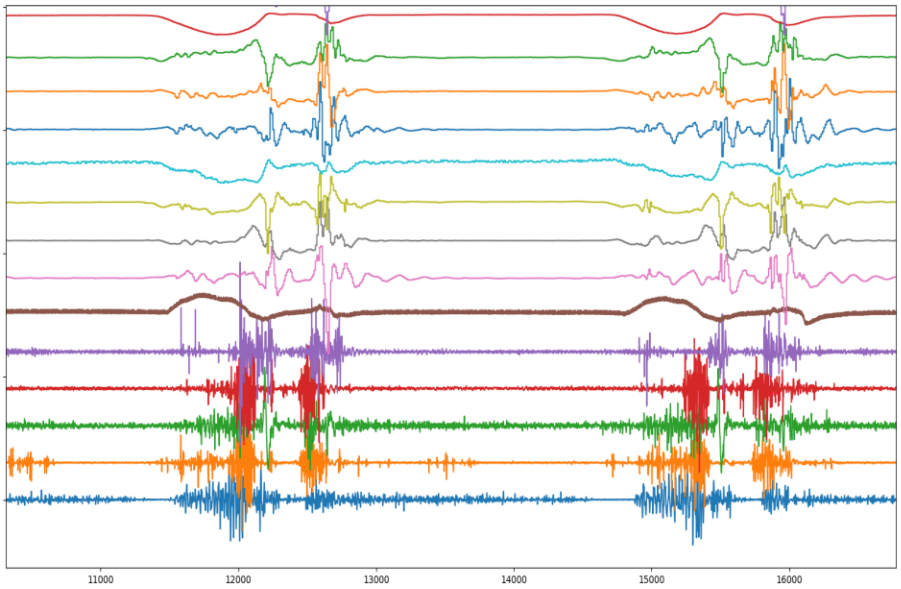


walk: all sensors, all recording sessions

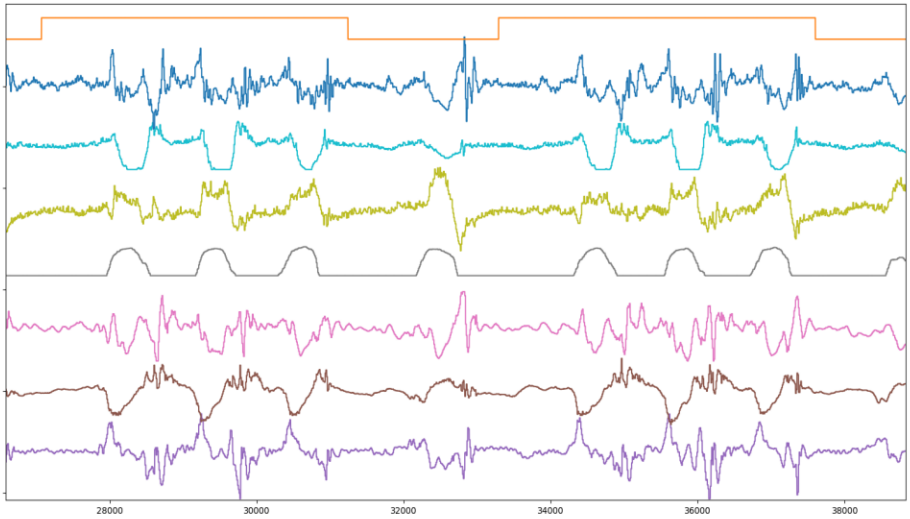


walk: all sensors, three gait cycles

Figure B.3 – Exemplar of the sensor data documentation: Page 3.



jump-two-leg: accelerometers + acoustic + EMG + gonio, twice



walk-curve-left: gyroscopes + force, two gait cycles

Figure B.4 – Exemplar of the sensor data documentation: Page 4.

Bibliography

Abraham, A. (2005). Artificial neural networks. *Handbook of measuring system design*.

ACC (2021). Website (accessed October 15, 2021). `https://biosignalsplux.com/products/sensors/accelerometer.html`.

Advene (2021). Website (accessed October 15, 2021). `https://advene.org/`.

Akay, M. (2000). *Nonlinear biomedical signal processing*. Wiley Online Library.

Ali, A. and Aggarwal, J. (2001). Segmentation and recognition of continuous human activity. In *Proceedings IEEE Workshop on Detection and Recognition of Events in Video*, pages 28–35. IEEE.

Ames, C. (1989). The markov process as a compositional model: A survey and tutorial. *Leonardo*, 22(2):175–187.

Amma, C., Gehrig, D., and Schultz, T. (2010). Airwriting recognition using wearable motion sensors. In *First Augmented Human International Conference*, page 10. ACM.

Ancans, A., Rozentals, A., Nesenbergs, K., and Greitans, M. (2017). Inertial sensors and muscle electrical signals in human-computer interaction. In *ICTA 2017 — 6th International Conference on Information and Communication Technology and Accessibility*, pages 1–6. IEEE.

Arifoglu, D. and Bouchachia, A. (2017). Activity recognition and abnormal behaviour detection with recurrent neural networks. *Procedia Computer Science*, 110:86–93.

Arous, M. A. B., Dunbar, M., Arfaoui, S., Mitiche, A., Ouakrim, Y., Fuentes, A., Richardson, G., and Mezghani, N. (2018). Knee kinematics feature selection for surgical and nonsurgical arthroplasty candidate characterization. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 3: BIOSIGNALS*, pages 176–181.

Arthrokinemat (2021). Website (accessed October 15, 2021). https://www.uni-bremen.de/en/csl/projects/past-projects/arthrokinemat.

Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Bakis, R. (1976). Continuous speech recognition via centisecond acoustic states. *The Journal of the Acoustical Society of America*, 59(S1):S97–S97.

Bao, L. and Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In *PERCOM 2014 — IEEE International Conference on Pervasive Computing and Communications*, pages 1–17. Springer.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). TSFEL: Time series feature extraction library. *SoftwareX*, 11:100456.

Barbič, J., Safonova, A., Pan, J.-Y., Faloutsos, C., Hodgins, J. K., and Pollard, N. S. (2004). Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface*, pages 185–194. Citeseer.

Bauerfeind-GenuTrain (2021). Website (accessed October 15, 2021). https://www.bauerfeind-group.com/en/products/supports-and-orthoses/knee-hip-thigh/details/product/genutrain.

BBDC (2019). Website (accessed October 15, 2021). https://bbdc.csl.uni-bremen.de/index.php/2019.

Beddiar, D. R., Nini, B., Sabokrou, M., and Hadid, A. (2020). Vision-based human activity recognition: A survey. *Multimedia Tools and Applications*, 79(41):30509–30555.

Bellman, R., Corporation, R., and Collection, K. M. R. (1957). *Dynamic Programming*. Rand Corporation research study. Princeton University Press.

biosignalsplux (2021). Website (accessed October 15, 2021). https://biosignalsplux.com/.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Bo, A. P. L., Hayashibe, M., and Poignet, P. (2011). Joint angle estimation in rehabilitation with inertial sensors and its integration with kinect. In *EMBC 2011 — 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3479–3483. IEEE.

BREMEN.AI (2021). Website (accessed October 15, 2021). `https://bremen.ai/events/`.

Carbonaro, N., Dalle Mura, G., Lorussi, F., Paradiso, R., De Rossi, D., and Tognetti, A. (2014). Exploiting wearable goniometer technology for motion sensing gloves. *IEEE journal of biomedical and health informatics*, 18(6):1788–1795.

Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Tröster, G., Millán, J. d. R., and Roggen, D. (2013). The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033–2042.

Davis, J. H. (2002). State space realizations. *Foundations of Deterministic and Stochastic Control*, pages 1–69.

De Leonardi, G., Rosati, S., Balestra, G., Agostini, V., Panero, E., Gastaldi, L., and Knaflitz, M. (2018). Human activity recognition by wearable sensors: Comparison of different classifiers for real-time applications. In *MeMeA 2018 — IEEE International Symposium on Medical Measurements and Applications*, pages 1–6. IEEE.

Demrozi, F., Pravadelli, G., Bihorac, A., and Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey. 1(1).

Deng, Z., Vahdat, A., Hu, H., and Mori, G. (2016). Structure inference machines: Recurrent neural networks for analyzing relations in group activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4772–4781.

Dey, N. and Ashour, A. (2016). *Classification and clustering in biomedical signal processing*. IGI global Hershey.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10):78.

Ejupi, A., Gschwind, Y. J., Valenzuela, T., Lord, S. R., and Delbaere, K. (2016). A kinect and inertial sensor-based system for the self-assessment of fall risk: A home-based study in older people. *HCI 2016 — 18th International Conference on Human-Computer Interaction*, 31(3-4):261–293.

Eliasson, O. (2017). Monitoring of doors, door handles and windows using inertial sensors.

EMG (2021). Website (accessed October 15, 2021). https://biosignalsplux.com/products/sensors/electromyography.html/.

Figueira, C., Matias, R., and Gamboa, H. (2016). Body location independent activity monitoring. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 4: BIOSIGNALS*, pages 190–197. INSTICC, SciTePress.

Fischer, C., Sukumar, P. T., and Hazas, M. (2012). Tutorial: Implementing a pedestrian tracker using inertial sensors. *IEEE pervasive computing*, 12(2):17–27.

Fleischer, C. and Reinicke, C. (2005). Predicting the intended motion with EMG signals for an exoskeleton orthosis controller. In *IROS 2005 — IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2029–2034.

FSR (2021). Website (accessed October 15, 2021). https://biosignalsplux.com/products/sensors/pressure.html.

Gehrig, D. (2015). *Automatic Recognition of Concurrent and Coupled Human Motion Sequences*. PhD thesis, Karlsruher Institut für Technologie.

Goniometer (2021). Website (accessed October 15, 2021). https://biosignalsplux.com/products/sensors/goniometer.html.

Graupe, D. (2013). *Principles of artificial neural networks*, volume 7. World Scientific.

Ground-Cable (2021). Website (accessed October 15, 2021). https://plux.info/accessories/472-ground-cable.html.

Groves, P. D. (2015). Navigation using inertial sensors [tutorial]. *IEEE Aerospace and Electronic Systems Magazine*, 30(2):42–69.

Guenterberg, E., Ostadabbas, S., Ghasemzadeh, H., and Jafari, R. (2009). An automatic segmentation technique in body sensor networks based on signal energy. In *Proceedings of the Fourth International Conference on Body Area Networks*. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering.

Gurvits, L. and Ledoux, J. (2005). Markov property for a function of a markov chain: A linear algebra approach. *Linear Algebra and its Applications*, 404:85–117.

Ha, S., Yun, J.-M., and Choi, S. (2015). Multi-modal convolutional neural networks for activity recognition. In *SMC 2015 — IEEE International conference on systems, man, and cybernetics*, pages 3017–3022. IEEE.

Han, S.-H., Ahn, E.-J., Ryu, M.-H., and Kim, J.-N. (2019). Natural hand gesture recognition with an electronic textile goniometer. *Sensors Mater*, 31:1387–1395.

Handheld-Switch (2021). Website (accessed October 15, 2021). `https://plux.info/actuators/325-trigger.html`.

Hartmann, Y. (2020). Feature selection for multimodal human activity recognition. Master's thesis, Universität Bremen.

Hartmann, Y., Liu, H., and Schultz, T. (2020). Feature space reduction for multimodal human activity recognition. In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 4: BIOSIGNALS*, pages 135–140. INSTICC, SciTePress.

Hartmann, Y., Liu, H., and Schultz, T. (2021). Feature space reduction for human activity recognition based on multi-channel biosignals. In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and TechnologiesS*, pages 215–222. INSTICC, SciTePress.

Hassoun, M. H. et al. (1995). *Fundamentals of artificial neural networks*. MIT press.

Hawkes, D. and Qu, Y. (1985). *The Songs of the South: An Ancient Chinese Anthology of Poems by Qu Yuan and Other Poets*. Penguin Books London.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *CVPR 2016 — IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Hellmers, S., Kromke, T., Dasenbrock, L., Heinks, A., Bauer, J. M., Hein, A., and Fudickar, S. (2018). Stair climb power measurements via inertial measurement units.

Hintermüller, C. (2016). *Advanced Biosignal Processing and Diagnostic Methods*. BoD–Books on Demand.

Hopfield, J. J. (1988). Artificial neural networks. *IEEE Circuits and Devices Magazine*, 4(5):3–10.

Hu, B., Rouse, E., and Hargrove, L. (2018). Benchmark datasets for bilateral lower-limb neuromechanical signals from wearable sensors during unassisted locomotion in able-bodied individuals. *Frontiers in Robotics and AI*, 5:14.

Inoue, M., Inoue, S., and Nishida, T. (2018). Deep recurrent neural network for mobile human activity recognition with high throughput. *Artificial Life and Robotics*, 23(2):173–185.

Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3):31–44.

Janke, M. and Diener, L. (2017). EMG-to-Speech: Direct generation of speech from facial electromyographic signals. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 25(12):2375–2385.

Jou, S.-C. S., Schultz, T., and Waibel, A. (2007). Continuous electromyographic speech recognition with a multi-stream decoding architecture. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Jung, Y. and Cha, B. (2010). Gesture recognition based on motion inertial sensors for ubiquitous interactive game contents. *IETE Technical Review*, 27(2):158–166.

Kahol, K., Tripathi, P., Panchanathan, S., and Rikakis, T. (2003). Gesture segmentation in complex motion sequences. In *Proceedings of 2003 International Conference on Image Processing*, volume 2, pages II–105. IEEE.

Ke, S.-R., Thuc, H. L. U., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H., and Choi, K.-H. (2013). A review on video-based human activity recognition. *Computers*, 2(2):88–131.

Keshavarzian, A., Sharifian, S., and Seyedin, S. (2019). Modified deep residual network architecture deployed on serverless framework of iot platform based on human activity recognition application. *Future Generation Computer Systems*, 101:14–28.

Kim, S., Kim, J., and Suh, D. (2019). Game controller position tracking using a2c machine learning on inertial sensors. In *GEM 2019 — IEEE Games, Entertainment, Media Conference*, pages 1–6. IEEE.

Kwapisz, J. R., Weiss, G. M., and Moore, S. A. (2010). Activity recognition using cell phone accelerometers. In *Proceedings of the 4th International Workshop on Knowledge Discovery from Sensor Data*, pages 10–18.

Kwon, Y., Kang, K., and Bae, C. (2015). Analysis and evaluation of smartphone-based human activity recognition using a neural network approach. In *IJCNN 2015 — International Joint Conference on Neural Networks*, pages 1–5. IEEE.

Lara, O. D. and Labrador, M. A. (2012). A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials*, 15(3):1192–1209.

Lee, S.-M., Yoon, S. M., and Cho, H. (2017). Human activity recognition from accelerometer data using convolutional neural network. In *BIGCOMP 2017 — IEEE International Conference on Big Data and Smart Computing*, pages 131–134. IEEE.

Li, F., Shirahama, K., Nisar, M. A., Köping, L., and Grzegorzek, M. (2018). Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors*, 18(2):1–22.

Li, Y. and Li, L. (2009). A novel split and merge EM algorithm for gaussian mixture model. In *2009 ICNC — 5th International Conference on Natural Computation*, volume 6, pages 479–483. IEEE.

Liang, H., Bronzino, J. D., and Peterson, D. R. (2012). *Biosignal processing: Principles and practices*. CRC Press.

Liu, H. (2019). Tutorial: From offline towards real-time: a wearable har system using biosensors integrated into a knee bandage. In Tutorial Section of *Biostec 2019 — 12th International Joint Conference on Biomedical Engineering Systems and Technologies*.

Liu, H., Hartmann, Y., and Schultz, T. (2021). Motion Units: Generalized sequence modeling of human activities for sensor-based activity recognition. In *EUSIPCO 2021 — 29th European Signal Processing Conference*. IEEE.

Liu, H. and Schultz, T. (2018). ASK: A framework for data acquisition and activity recognition. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 3: BIOSIGNALS*, pages 262–268. INSTICC, SciTePress.

Liu, H. and Schultz, T. (2019). A wearable real-time human activity recognition system using biosensors integrated into a knee bandage. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 1: BIODEVICES*, pages 47–55. INSTICC, SciTePress.

Long, J., Sun, W., Yang, Z., and Raymond, O. I. (2019). Asymmetric residual neural network for accurate human activity recognition. *Information*, 10(6):203.

Loose, H., Tetzlaff, L., and Bolmgren, J. L. (2020). A public dataset of overground and treadmill walking in healthy individuals captured by wear-

able imu and semg sensors. In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 4: BIOSIGNALS*, pages 164–171. INSTICC, SciTePress.

Lukowicz, P., Ward, J. A., Junker, H., Stäger, M., Tröster, G., Atrash, A., and Starner, T. (2004). Recognizing workshop activity using body worn microphones and accelerometers. In *In Pervasive Computing*, pages 18–32.

LUX (2021). Website (accessed October 15, 2021). `https://plux.info/sensors/296-light-lux.html`.

Mai, L. (2019). Automatische segmentierung von bewegungsdaten eines biosignalbasierten, in einer kniebandage integrierten har systems. Bachelor's thesis, Universität Bremen.

Mason, C., Gadzicki, K., Meier, M., Ahrens, F., Kluss, T., Maldonado, J., Putze, F., Fehr, T., Zetzsche, C., Herrmann, M., Schill, K., and Schultz, T. (2020). From human to robot everyday activity. In *IROS 2020 — IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, USA.

Mason, C., Meier, M., Ahrens, F., Fehr, T., Herrmann, M., Putze, F., and Schultz, T. (2018). Human activities data collection and labeling using a think-aloud protocol in a table setting scenario. In *IROS 2018 — IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Mathie, M. J., Coster, A. C. F., Lovell, N., and Celler, B. (2003). Detection of daily physical activities using a triaxial accelerometer. In *Medical and Biological Engineering and Computing*. 41(3):296—301.

Mehrotra, K., Mohan, C. K., and Ranka, S. (1997). *Elements of artificial neural networks*. MIT press.

Meier, M., Mason, C., Porzel, R., Putze, F., and Schultz, T. (2018). Synchronized multimodal recording of a table setting dataset. IROS 2018 — IEEE/RSJ International Conference on Intelligent Robots and Systems.

Merletti, R. and Di Torino, P. (1999). Standards for reporting EMG data. *J Electromyogr Kinesiol*, 9(1):3–4.

MetaMotionR (2021). Website (accessed October 15, 2021). `https://mbientlab.com/metamotionr/`.

Mezghani, N., Fuentes, A., Gaudreault, N., Mitiche, A., Aissaoui, R., Hagmeister, N., and De Guise, J. A. (2013). Identification of knee frontal plane

kinematic patterns in normal gait by principal component analysis. *Journal of Mechanics in Medicine and Biology*, 13(03):1350026.

Micucci, D., Mobilio, M., and Napoletano, P. (2017). UniMiB SHAR: A dataset for human activity recognition using acceleration data from smartphones. *Applied Sciences*, 7(10):1101.

Miikkulainen, R., Liang, J. Z., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzyan, A., Duffy, N., and Hodjat, B. (2017). Evolving deep neural networks. *CoRR*, abs/1703.00548.

Mu, C., Xie, J., Yan, W., Liu, T., and Li, P. (2016). A fast recognition algorithm for suspicious behavior in high definition videos. *Multimedia Systems*, 22(3):275–285.

Multi-Sync-Splitter (2021). Website (accessed October 15, 2021). `https://plux.info/cables/300-multi-sync-splitter-820201704.html`.

Murad, A. and Pyun, J.-Y. (2017). Deep recurrent neural networks for human activity recognition. *Sensors*, 17(11):2556.

Naït-Ali, A. (2009). *Advanced biosignal processing*. Springer Science & Business Media.

Nguyen, L. (2016). Continuous observation hidden markov model. 44(6):65–149.

NumPy (2021). Website (accessed October 15, 2021). `numpy.org`.

Nurhanim, K., Elamvazuthi, I., Izhar, L., and Ganesan, T. (2017). Classification of human activity based on smartphone inertial sensor using support vector machine. In *ROMA 2017 — 3rd IEEE International Symposium in Robotics and Manufacturing Automation*, pages 1–5. IEEE.

Oniga, S. and Sütő, J. (2014). Human activity recognition using neural networks. In *Proceedings of the 15th International Carpathian Control Conference*, pages 403–406. IEEE.

Ordóñez, F. J. and Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1):115.

Palyafári, R. (2015). Continuous activity recognition for an intelligent knee orthosis; an out-of-lab study. Master's thesis, Karlsruher Institut für Technologie.

Park, S. K. and Suh, Y. S. (2010). A zero velocity detection algorithm using inertial sensors for pedestrian navigation systems. *Sensors*, 10(10):9163–9178.

Peeters, G. (2004). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. *CUIDADO Ist Project Report*, 54(0):1–25.

Phinyomark, A., Hirunviriya, S., Limsakul, C., and Phukpattaranont, P. (2010). Evaluation of EMG feature extraction for hand movement recognition based on euclidean distance and standard deviation. In *ECTI-CON 2010 — 7th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technolog*, pages 856–860. IEEE.

Priddy, K. L. and Keller, P. E. (2005). *Artificial neural networks: An introduction*, volume 68. SPIE press.

PZT (2021). Website (accessed October 15, 2021). `https://biosignalsplux.com/products/sensors/respiration-piezo.html`.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, volume 77, pages 257–286.

Rabiner, L. R. and Juang, B. H. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–16.

Raya, R., Rocon, E., Gallego, J. A., Ceres, R., and Pons, J. L. (2012). A robust kalman algorithm to facilitate human-computer interaction for people with cerebral palsy, using a new interface based on inertial sensors. *Sensors*, 12(3):3049–3067.

Rebelo, D., Amma, C., Gamboa, H., and Schultz, T. (2013). Human activity recognition for an intelligent knee orthosis. In *Proceedings of the 6th International Conference on Bio-inspired Systems and Signal Processing - BIOSIGNALS*, pages 368–371.

Reiss, A. and Stricker, D. (2012a). Creating and benchmarking a new dataset for physical activity monitoring. In *Proceedings of the 5th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 1–8.

Reiss, A. and Stricker, D. (2012b). Introducing a new benchmarked dataset for activity monitoring. In *ISWC 2012 — 16th International Symposium on Wearable Computers*, pages 108–109. IEEE.

Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkl, G., Ferscha, A., et al. (2010). Collecting complex activity datasets in highly rich networked sensor environments. In *INSS 2010 — 7th International Conference on Networked Sensing Systems*, pages 233–240. IEEE.

Ronao, C. A. and Cho, S.-B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications*, 59:235–244.

Ronaoo, C. A. and Cho, S.-B. (2015). Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors. *Journal of the Korean Information Science Society*, pages 858–860.

Rowe, P., Myles, C. M., Walker, C., and Nutton, R. (2000). Knee joint kinematics in gait and other functional activities measured using flexible electrogoniometry: How much knee motion is sufficient for normal daily life? *Gait & posture*, 12(2):143—-155.

Rubenson, J., Heliams, D. B., Lloyd, D. G., and Fournier, P. A. (2004). Gait selection in the ostrich: Mechanical and metabolic characteristics of walking and running with and without an aerial phase. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1543):1091–1099.

Santos, S., Folgado, D., and Gamboa, H. (2020). Upper-body movements: Precise tracking of human motion using inertial sensors.

Saritha, C., Sukanya, V., and Murthy, Y. N. (2008). ECG signal analysis using wavelet transforms. *Bulg. J. Phys*, 35(1):68–77.

Schultz, T. (2019). Biosignale und Benutzerschnittstellen, lecture notes: Einführung.

Siddiqi, M. H., Ali, R., Rana, M., Hong, E.-K., Kim, E. S., Lee, S., et al. (2014). Video-based human activity recognition using multilevel wavelet decomposition and stepwise linear discriminant analysis. *Sensors*, 14(4):6370–6392.

Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., and Holzinger, A. (2017). Human activity recognition using recurrent neural networks. In *CD-MAKE 2017 — International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, pages 267–274. Springer.

Sousa, W., Souto, E., Rodrigres, J., Sadarc, P., Jalali, R., and El-Khatib, K. (2017). A comparative analysis of the impact of features on human activity recognition with smartphone sensors. In *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*, pages 397–404.

Sousa Lima, W., Souto, E., El-Khatib, K., Jalali, R., and Gama, J. (2019). Human activity recognition using inertial sensors in a smartphone: An overview. *Sensors*, 19(14):3213.

Sprager, S. and Juric, M. B. (2015). Inertial sensor-based gait recognition: A review. *Sensors*, 15(9):22089–22127.

Stetter, B., Krafft, F., Ringhof, S., Sell, S., and Stein, T. (2019a). Assessing knee joint forces using wearable sensors and machine learning techniques. In *Proceedings der DVS-Jahrestagung Biomechanik 2019*, pages 55–56. Universität Konstanz.

Stetter, B. J., Krafft, F. C., Ringhof, S., Gruber, R., Sell, S., and Stein, T. (2018). Estimation of the knee joint load in sport-specific movements using wearable sensors. In *SPINFORTEC 2018 — 12th Symposium der Sektion Sportinformatik und Sporttechnologie der Deutschen Vereinigung*.

Stetter, B. J., Krafft, F. C., Ringhof, S., Stein, T., and Sell, S. (2020). A machine learning and wearable sensor based approach to estimate external knee flexion and adduction moments during various locomotion tasks. *Frontiers in Bioengineering and Biotechnology*, 8:article: 9.

Stetter, B. J., Ringhof, S., Krafft, F., Sell, S., and Stein, T. (2019b). Estimation of knee joint forces in sport movements using wearable sensors and machine learning. *Sensors*, 19(17):article: 3690.

Sutherland, D. H. (2002). The evolution of clinical gait analysis: Part II kinematics. *Gait & Posture*, 16(2):159–179.

Synchronization-Cable (2021). Website (accessed October 15, 2021). `https://plux.info/cables/231-synchronization-cable.html`.

Synchronization-Kit (2021). Website (accessed October 15, 2021). `https://plux.info/sensors/322-synchronization-kit.html`.

Sztyler, T. and Stuckenschmidt, H. (2016). On-body localization of wearable devices: An investigation of position-aware activity recognition. In *PerCom 2016 — 14th IEEE International Conference on Pervasive Computing and Communications*, pages 1–9. IEEE Computer Society.

Takeda, S., Lago, P., Okita, T., and Inoue, S. (2019). Reduction of marker-body matching work in activity recognition using motion capture. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pages 835–842.

Teague, C. N., Hersek, S., Toreyin, H., Millard-Stafford, M. L., Jones, M. L., Kogler, G. F., Sawka, M. N., and Inan, O. T. (2016). Novel methods for sensing acoustical emissions from the knee for wearable joint health assessment. *IEEE Transactions on Biomedical Engineering*, 63(8):1581–1590.

Telaar, D., Wand, M., Gehrig, D., Putze, F., Amma, C., Heger, D., Vu, N. T., Erhardt, M., Schlippe, T., Janke, M., et al. (2014). Biokit—real-time decoder for biosignal processing. In *INTERSPEECH 2014 — 15th Annual Conference of the International Speech Communication Association*.

Theis, F. J. and Meyer-Bäse, A. (2010). *Biomedical signal analysis: Contemporary methods and applications*. MIT Press.

Tolstikov, A., Hong, X., Biswas, J., Nugent, C., Chen, L., and Parente, G. (2011). Comparison of fusion methods based on DST and DBN in human activity recognition. *Journal of Control Theory and Applications*, 9(1):18–27.

Tuncer, T., Ertam, F., Dogan, S., Aydemir, E., and Pławiak, P. (2020). Ensemble residual network-based gender and activity recognition method with signals. *The Journal of Supercomputing*, 76(3):2119–2138.

Uddin, M. Z., Lee, J., and Kim, T.-S. (2008). Independent component feature-based human activity recognition via linear discriminant analysis and hidden markov model. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5168–5171. IEEE.

Ueda, N., Nakano, R., Ghahramani, Z., and Hinton, G. E. (2000). Split and merge EM algorithm for improving gaussian mixture density estimates. *Journal of VLSI signal processing systems for signal, image and video technology*, 26(1-2):133–140.

Urban, T. (2019). Entwicklung einer mobilen anwendung für echtzeit-visualisierung und archivierung von mehrkanaliger biosignalaufnahme mit bluetooth. Bachelor's thesis, Universität Bremen.

Van Kasteren, T., Englebienne, G., and Kröse, B. J. (2010). An activity monitoring system for elderly care using generative and discriminative models. *Personal and ubiquitous computing*, 14(6):489–498.

Veth, M. and Raquet, J. (2006). Fusion of low-cost imagining and inertial sensors for navigation. In *Proceedings of the 19th International Technical*

Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2006), pages 1093–1103.

Veth, M. J. (2006). Fusion of imaging and inertial sensors for navigation. Technical report, Air Force Institute of Technology — Wright Patterson Air Force Base Ohio, School of Engineering and Management.

Vicon (2021). Website (accessed October 15, 2021). `https://www.vicon.com/`.

Vrigkas, M., Nikou, C., and Kakadiaris, I. A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2:28.

Wand, M. and Schultz, T. (2014). Pattern learning with deep neural networks in EMG-based speech recognition. In *EMBC 2014 — 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.

Wang, A., Chen, G., Yang, J., Zhao, S., and Chang, C.-Y. (2016a). A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sensors Journal*, 16(11):4566–4578.

Wang, H., Oneata, D., Verbeek, J., and Schmid, C. (2016b). A robust and efficient video representation for action recognition. *International journal of computer vision*, 119(3):219–238.

Wang, J., Chen, Y., Hao, S., Peng, X., and Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, 119:3–11.

Wang, L. (2020). *Computer Methods and Programs in Biomedical Signal and Image Processing*. BoD–Books on Demand.

Wang, X., Tarrío, P., Metola, E., Bernardos, A. M., and Casar, J. R. (2012). Gesture recognition using mobile phone's inertial sensors. In *Distributed Computing and Artificial Intelligence*, pages 173–184. Springer.

Watson, B. (1994). Selected poems of Su Tung-p'o. *Port Townsend, Wa: Copper canyon*.

Weinland, D., Ronfard, R., and Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. *Computer vision and image understanding*, 115(2):224–241.

Whittle, M. W. (1996). Clinical gait analysis: A review. *Human movement science*, 15(3):369–387.

Whittle, M. W. (2014). *Gait analysis: An introduction*. Butterworth-Heinemann.

Woodman, O. J. (2007). An introduction to inertial navigation. Technical report, University of Cambridge, Computer Laboratory.

Wu, C.-H., Chang, Y.-T., and Tseng, Y.-C. (2010). Multi-screen cyber-physical video game: An integration with body-area inertial sensor networks. In *PERCOM Workshops 2010 — 8th IEEE International Conference on Pervasive Computing and Communications Workshops*, pages 832–834. IEEE.

Xue, T. and Liu, H. (2021). Hidden Markov Model and its application in human activity recognition and fall detection: A review. In *CSPS 2021 — 10th International Conference on Communications, Signal Processing, and Systems*. forthcoming.

Yang, J., Lee, J., and Choi, J. (2011). Activity recognition based on RFID object usage for smart mobile devices. *Journal of Computer Science and Technology*, 26(2):239–246.

Yang, J., Nguyen, M. N., San, P. P., Li, X., and Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *IJCAI*, volume 15, pages 3995–4001. Buenos Aires, Argentina.

Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., and Zhang, J. (2014). Convolutional neural networks for human activity recognition using mobile sensors. In *MOBICASE 2014 — 6th International Conference on Mobile Computing, Applications and Services*, pages 197–205. IEEE.

Zhang, X., Chen, X., Li, Y., Lantz, V., Wang, K., and Yang, J. (2011). A framework for hand gesture recognition based on accelerometer and EMG sensors. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6):1064–1076.

Zhang, Z., Chen, C., Sun, J., and Chan, K. L. (2003). EM algorithms for gaussian mixtures with split-and-merge operation. *Pattern recognition*, 36(9):1973–1983.

Zok, M. (2014). Inertial sensors are changing the games. In *ISISS 2014 — International Symposium on Inertial Sensors and Systems*, pages 1–3. IEEE.

Addendum

List of Supervised Student Theses

The following table presents a list of all student theses (Bachelorarbeit or Masterarbeit) that were supervised by the author. All of these were completed under the primary supervision of Prof. Dr.-Ing. Tanja Schultz.

Student	Year	Thesis Title
Yale Hartmann	2019	Implementation and Optimisation of a Human Activity Recognition System using Sensors integrated into a Knee Bandage (Bachelorarbeit)
Timo Urban	2019	Entwicklung einer mobilen Anwendung für Echtzeit-Visualisierung und Archivierung von mehrkanaliger Biosignalaufnahme mit Bluetooth (Bachelorarbeit)
Lennard Mai	2019	Automatische Segmentierung von Bewegungsdaten eines biosignalbasierten, in einer Kniebandage integrierten HAR Systems (Bachelorarbeit)
Yale Hartmann	2020	Feature Selection for Multimodal Human Activity Recognition (Masterarbeit)
Kilian Lüdemann	2021	Sturzerkennung in Infrarotvideos mit Hilfe von Rekurrenten Neuronalen Netzen (Bachelorarbeit)
Steffen Lahrberg	2021	Direction Distinguishment in Wearable Sensor-Based Human Activity Recognition Using Cross-Channel Features (Bachelorarbeit)

List of Publications

The following bibliography lists all publications of which the author of the dissertation is the first author or co-author.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). TSFEL: Time series feature extraction library. *SoftwareX*, 11:100456.

Folgado, D., Barandas, M., Antunes, M., Nunes, M. L., Liu, H., Hartmann, Y., Schultz, T., and Gamboa, H. (2021). TSSEARCH: Time series subsequence search library. *SoftwareX*. forthcoming.

Hartmann, Y., Liu, H., Lahrberg, S., and Schultz, T. (2022). Interpretable high-level features for human activity recognition. In *BIOSIGNALS 2022 — 15th International Conference on Bio-inspired Systems and Signal Processing*. INSTICC, SciTePress. forthcoming.

Hartmann, Y., Liu, H., and Schultz, T. (2020). Feature space reduction for multimodal human activity recognition. In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 4: BIOSIGNALS*, pages 135–140. INSTICC, SciTePress.

Hartmann, Y., Liu, H., and Schultz, T. (2021). Feature space reduction for human activity recognition based on multi-channel biosignals. In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and TechnologiesS*, pages 215–222. INSTICC, SciTePress.

Liu, H., Hartmann, Y., and Schultz, T. (2021a). CSL-SHARE: A multimodal wearable sensor-based human activity dataset. *Frontiers in Computer Science*, 3:90.

Liu, H., Hartmann, Y., and Schultz, T. (2021b). Motion Units: Generalized sequence modeling of human activities for sensor-based activity recognition. In *EUSIPCO 2021 — 29th European Signal Processing Conference*. IEEE.

Liu, H. and Schultz, T. (2018). ASK: A framework for data acquisition and activity recognition. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies — Volume 3: BIOSIGNALS*, pages 262–268. INSTICC, SciTePress.

Liu, H. and Schultz, T. (2019). A wearable real-time human activity recognition system using biosensors integrated into a knee bandage. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Sys-*

tems and Technologies — Volume 1: BIODEVICES, pages 47–55. INSTICC, SciTePress (Best Student Paper).

Liu, H. and Wang, X. (2011). Capacity of cooperative ad hoc networks with heterogeneous traffic patterns. In *ICC 2011 — IEEE International Conference on Communications*, pages 1–5. IEEE.

Weiner, J., Diener, L., Stelter, S., Externest, E., Kühl, S., Herff, C., Putze, F., Schulze, T., Salous, M., Liu, H., Küster, D., and Schultz, T. (2017). Bremen Big Data Challenge 2017: Predicting university cafeteria load. In Kern-Isberner, G., Fürnkranz, J., and Thimm, M., editors, *KI 2017 — Advances in Artificial Intelligence: 40th Annual German Conference on AI, Dortmund, Germany, Proceedings*. Springer International Publishing.

Xue, T. and Liu, H. (2021). Hidden Markov Model and its application in human activity recognition and fall detection: A review. In *CSPS 2021 — 10th International Conference on Communications, Signal Processing, and Systems*. forthcoming.

感知

劉輝

你看到山川多峻峭
因為眼睛傳輸信號
你聽到鳥鳴多繚繞
因為耳朵也是信道

你說，你笑
不僅人類可以感受到
你玩，你鬧
機器的陪伴也很可靠

你的肌膚和你的大腦
你的溫度和你的心跳
我們感知這一切奇妙
並不為認出你的容貌
只是想讓你變得重要
只是想讓生活更美好

Perception

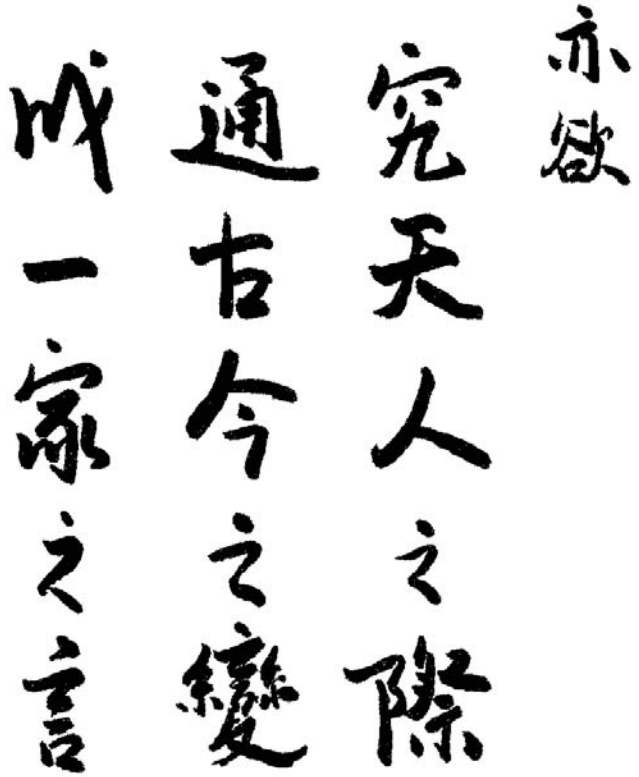
for Cognitive Systems Lab

Hui Liu

You see all things recover in spring
'cause your eyes perceive signals
You hear how beautiful birds sing
'cause your ears are also channels

You laugh, you cry
Not only humans can feel
You talk, you sigh
Machines always keep their zeal

Your brain, your heat
Your muscle and heartbeat
We sense the trace
But never recog' your face
Just to make life light
Just to make you bright



"I also intended

to investigate the correlation between nature and human being;

to review the ancient and modern evolution extensively;

to elucidate my own system of analysis."

Sima Qian (145 BC — ?), *Records of the Grand Historian.*

人體動作識別中的生物訊號處理和動作建模 劉琦

Biosignal Processing and Activity Modeling for Human Activity Recognition