Integrating Passive Ubiquitous Surfaces into Human-Computer Interaction

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
(abgekürzt: Dr. rer. nat.)
genehmigte Dissertation

von Herrn
M.Sc. Maximilian Schrapel
geboren am 12. April 1989
in Magdeburg, Deutschland

2022

Integrating Passive Ubiquitous Surfaces into Human-Computer Interaction
Dissertation
Maximilian Schrapel

Referent:
Prof. Dr. Michael Rohs
Gottfried Wilhelm Leibniz Universität Hannover

Korreferent:
Prof. Dr. Johannes Schöning
Universität St.Gallen

Tag der Promotion: 21. Oktober 2022

Gottfried Wilhelm Leibniz Universität Hannover
Institut für Mensch-Computer-Interaktion
Fakultät für Elektrotechnik und Informatik
Appelstr. 9A
30167 Hannover

Abstract

Computers are ubiquitous. They simplify everyday tasks and make globally connected information and communication available anytime and anywhere. However, we are often still confined to a computer screen. One approach to expand the interaction space of computers is to embed systems in objects and surfaces. However, this approach has limited scalability, which restricts the ubiquitous use of technology.

This dissertation investigates how ubiquitous, unmodified surfaces from the user's natural environment can be integrated into human-computer interactions. For this purpose, we introduce the term passive surface interactions. The aim is to enable a seamless transition to virtual worlds. Several prototypes are used to investigate and demonstrate how the properties of passive surfaces can contribute to identifying the user, the interaction, as well as the surface itself. Hereby, this work focuses on visual, acoustic and tactile features of a surface. Furthermore, we show how the perception of everyday passive surfaces can be influenced and how this can be used for interacting with computers.

It turns out that surface features can be used in a versatile way to partially substitute embedded systems in objects and surfaces. A key insight is that scratching sounds that are emitted during interaction contain features that are complementary to the motion and vibration signals that are produced during such an interaction. Audio and motion sensors can be combined to identify input events, the users themselves, as well as the surface. Moreover, visual features offer a wide range of applications, especially for heterogeneous color distributions of many different surfaces. We show that acoustic augmentations of surfaces can intuitively convey information and unnoticeably influence the user's behavior. Furthermore, acoustic surface augmentation contributes to the perception of a virtual environment. However, privacy concerns must be considered in the ubiquitous use of sensor technology. The study results regarding privacy show that context can be crucial in deciding whether and which interaction should be carried out.

Key words: Ubiquitous Computing, Surface Computing, Augmented Reality

Zusammenfassung

Computer sind allgegenwärtig. Sie vereinfachen alltägliche Aufgaben und stellen weltweit vernetzte Informationen sowie Kommunikation jederzeit und überall zur Verfügung. Allerdings sind wir oft noch an einen Computerbildschirm gebunden. Ein Ansatz zur Erweiterung des Interaktionsraums von Computern ist die Einbettung von Systemen in Objekte und Oberflächen. Dieser Ansatz ist jedoch nur begrenzt skalierbar, was den allgegenwärtigen Einsatz von Technologie einschränkt.

In dieser Dissertation wird erforscht, wie sich allgegenwärtige, unveränderte Oberflächen aus der natürlichen Umgebung des Nutzers in Mensch-Computer Interaktionen integrieren lassen. Hierfür führen wir den Begriff passive Oberflächen Interaktionen ein. Das Ziel hierbei ist einen nahtlosen Übergang zu virtuellen Welten zu ermöglichen. Es wird anhand mehrerer Prototypen untersucht und demonstriert, wie die Eigenschaften von passiven Oberflächen dazu beitragen können den Nutzer, seine Interaktion sowie die Oberfläche selbst zu erkennen. Diese Arbeit legt hierbei den Schwerpunkt auf visuelle, akustische und taktile Merkmale einer Oberfläche. Weiterhin zeigen wir wie sich die Wahrnehmung von alltäglichen, passiven Oberflächen beeinflussen lässt und wie das für die Interaktion mit Computern genutzt werden kann.

Es zeigt sich, dass Oberflächenmerkmale vielseitig eingesetzt werden können, um eingebettete Systeme in Objekten und Oberflächen teilweise zu ersetzen. Eine wichtige Erkenntnis ist, dass Kratzgeräusche, die während einer Interaktion erzeugt werden, Merkmale enthalten, die komplementär zu den Bewegungs- und Vibrationssignalen sind, die während einer solchen Interaktion erzeugt werden. Audio- und Bewegungssensoren können kombiniert werden um Eingaben, den Nutzer selbst, sowie die Oberfläche zu identifizieren. Weiterhin bieten visuelle Merkmale vielfältige Anwendungsmöglichkeiten, insbesondere bei heterogenen Farbverteilungen vieler verschiedener Oberflächen. Wir zeigen, dass akustische Augmentation von Oberflächen intuitiv Informationen übermitteln können und unbemerkt das Verhalten von Nutzern beeinflussen. Weiterhin trägt eine akustische Oberflächenaugmentation zur Wahrnehmung einer virtuellen Umgebung bei. Jedoch sind bei der ubiquitären Nutzung von Sensorik Datenschutzbedenken zu berücksichtigen. Die Ergebnisse der Studien in Bezug auf die Privatsphäre zeigen, dass der Kontext entscheidend sein kann, wenn es darum geht, ob und welche Interaktion durchgeführt werden soll.

Schlagwörter: Ubiquitäres Computing, Oberflächen Computing, Erweiterte Realität

Acknowledgments

Over the past few years as a doctoral student, I have met and worked with many amazing people who have inspired and supported me in many ways. They have motivated me to achieve my research goals, inspired me to try new ways of doing things, and supported me to do research at a high level. Their fruitful discussions, collaborations, and both professional and mental support have contributed to this dissertation. I would like to apologize to all the people who have not been mentioned here despite their contribution. Their support, discussions and ideas have not been forgotten and have been incorporated into this thesis. I would like to take this opportunity to express my deepest gratitude to all of them.

First of all, I would like to express my deepest thank to my supervisor Prof. Dr. Michael Rohs. Since the beginning of my work at the Human-Computer Interaction Group, he has placed his complete trust in me and gave me the opportunity to freely pursue my own research interests. Whenever I needed support, he was there to give advice and assistance. Michael's vast experience in HCI, which he always shared with me, the enormous efforts he made to support my research, and his excellent mentoring contributed greatly to my research success. I guess his level of support surpasses many other research groups in the field of HCI, for which I am very grateful.

Next, I would like to thank my second examiner Prof. Dr. Johannes Schöning. His research in surface computing inspired my research and thus contributed to the topic of my dissertation. In addition, he supported me in conducting a workshop, for which I am very grateful. I feel honored to have had such experienced researchers as examiners for my dissertation.

I would also like to thank my colleagues for their input and mental support. In particular, Tim Dünte, Jan Wolff, Kin Woon Yeow and Dennis Stanke as well as Matthias Becker and Jan Kassel contributed to my research through fruitful discussions and reliable teamwork. In addition to my supervised bachelor's and master's students, I would like to thank Benjamin Simon, Dennis Grannemann, Sergeij Löwen and Kamillo Ferry for their assistance as well as the numerous anonymous participants of my user studies who have tested the presented prototypes of this dissertation. My deepest gratitude is expressed for their honest feedback and interest in my research. Without their help this work would not have been possible. In addition, I would like to thank the secretary of the HCI Group Simone Vollmers for her support.

Over the years, I have worked with many inspiring researchers in Michael's group that I would like to mention here. First of all, I would like to thank the members of the project "Mobiler Mensch" and especially the organizers Prof. Dr. Kurt Schneider and Prof. Dr. Monika Sester. The project has allowed me to collaborate with many other researchers from different disciplines. To name a few, I would like to thank Oskar Wage, Udo Feuerhake, Christian Koetsier, Melanie Busch, Larissa Chazette, Kai Korte, Anne Finger and Lena Greinke for our joint work.

In addition, I have been fortunate to work with many researchers in the HCI community over the past years. I especially thank Jonathan Liebers and Stefan Schneegass for their collaboration on the skiing project at the Winter School 2020 which was later published at MUM'20. Furthermore, I was able to organize two workshops with many experienced researchers. In particular, Alexandra Voit, Johannes Schering, Shadan Sadeghian Borojeni and Philipp Wintersberger contributed to the successful organization of the workshops. Their input as well as the authors of the accepted papers and deep fruitful discussions on the workshop days have contributed to my research ideas.

Furthermore, the Leibniz University Hannover has given me the opportunity to present my research in industry. I would like to thank Prof. Dr. Wolfgang Nejdl for allowing me to exhibit my research at the "Marktplatz der Möglichkeiten". Furthermore, I thank Tobias Meyer and the team of "staring business" from the Leibniz University Hannover as well as Serge Enns and the team of "Nexster" from the University of Applied Sciences Hannover for their support. Both institutions opened up new opportunities for me to present my research to the industry. For example, I was fortunate to meet the CEO Thomas Bade of the "Institute for Universal Design", who graciously gave me the opportunity to present my research to a broader audience at the "Leitmesse Altenpflege 2021". In addition, I would like to thank Prof. Dr. Uwe Groth for supporting my project and the Pentelligence team from the leanlab 2021 Flo, Jonas, Nils and Henrik for their input.

Last but not least, I would like to deeply thank my family for their continuous support. In particular, my mother Annett Schrapel as well as my grandmother Renate Schrapel have supported me unconditionally my entire life to achieve my own goals. Likewise, Jörg Breitenstein and Kathrin Resch as well as Christian Resch always supported me to go my own way. Without all of your love, trust and support, this would not have been possible.

List of Contributing Publications and Supervised Theses

In the course of my time as a doctoral student I have supervised several Bachelor and Master theses. Some of the student theses contributed to this dissertation and related publications. The underlying publications were written in close cooperation with my doctoral advisor Michael Rohs. In addition, workshops were conducted and papers published in collaboration with other research groups. To acknowledge their contributed work and inspirations, the scientific "we" is used throughout this dissertation.

Publications

Several full and short papers at international conferences MuC, MUM, MobileHCI and CHI as well as a journal paper at i-com and a workshop at MobileHCI have contributed to this dissertation. In the following, the publications that contributed to this dissertation are listed chronologically.

- Schrapel, Maximilian & Stadler, Max-Ludwig & Rohs, Michael: 'Pentelligence: Combining Pen Tip Motion and Writing Sounds for Handwritten Digit Recognition'. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* New York, NY, USA: Association for Computing Machinery, 2018: pp. 1–11
- Schrapel, Maximilian & Herzog, Florian & Ryll, Steffen & Rohs, Michael: 'Watch My Painting: The Back of the Hand as a Drawing Space for Smartwatches'. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, 2020: pp. 1–10
- Schrapel, Maximilian & Liebers, Jonathan & Rohs, Michael & Schneegass, Stefan: 'Skiables: Towards a Wearable System Mounted on a Ski Boot for Measuring Slope Conditions'. *19th International Conference on Mobile and Ubiquitous Multimedia.* MUM 2020. Essen, Germany: Association for Computing Machinery, 2020: pp. 320–322
- Shadan Sadeghian & Maximilian Schrapel et al.: *Workshop on Designing Technologies for Future Forms of Sustainable Mobility. MobileHCI'20, Oldenburg, Germany.* 2020

- Schrapel, Maximilian & Schulz, Thilo & Rohs, Michael: 'Augmenting Public Bookcases to Support Book Sharing'. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '20. Oldenburg, Germany: Association for Computing Machinery, 2020
- Schrapel, Maximilian & Etgeton, Philipp & Rohs, Michael: 'SpectroPhone: Enabling Material Surface Sensing with Rear Camera and Flashlight LEDs'. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021
- Schrapel, Maximilian et al.: 'Sign H3re: Symbol and X-Mark Writer Identification Using Audio and Motion Data from a Digital Pen'. *Proceedings of Mensch Und Computer 2022*. MuC '22. Darmstadt, Germany: Association for Computing Machinery, 2022: pp. 209–218
- Schrapel, Maximilian and Happe, Janko and Rohs, Michael: 'EnvironZen: Immersive Soundscapes via Augmented Footstep Sounds in Urban Areas'. *i-com* (2022), vol. 21(2): pp. 219–237

Supervised Theses

The following Bachelor and Master Thesis as well as a student semester project under my supervision have contributed to this dissertation.

- Stadler, Max-Ludwig: 'Deep Learning zur Handschriftenerkennung mittels Audio- und Bewegungsdaten von Stiften'. Master Thesis. Leibniz Universität Hannover, 2017
- Nagel, Lukas: 'Using pen movement and audio data to verify signatures'. Bachelor Thesis. Leibniz Universität Hannover, 2018
- Grannemann, Dennis: 'Gesture Recognition with Audio and Motion Data from Digital Pens'. Bachelor Thesis. Leibniz Universität Hannover, 2018
- Schulz, Thilo: 'Entwicklung einer mobilen AR-Anwendungsunterstützung der visuellen Suche'. Master Thesis. Leibniz Universität Hannover, 2019
- Herzog, Florian: 'Implementation und Evaluation eines magnetischen Zeigegerätes für Smartwatches'. Bachelor Thesis. Leibniz Universität Hannover, 2019
- Ryll, Steffen: 'Interaction Techniques for Magnetic Pointing Devices on Smartwatches'. Bachelor Thesis. Leibniz Universität Hannover, 2020
- Bressgott, Timon: 'Strassenprofilschäztung für Fahrräder mit Audio- und Bewegungssignalen unter Verwendung Neuronaler Netze'. Master Thesis. Leibniz Universität Hannover, 2020
- Happe, Janko: 'Augmenting Footstep Sounds with Noise Cancelling Headphones in Public Places'. Master Thesis. Leibniz Universität Hannover, 2021
- Sado, Garros and Schmuhl, Julian and Schroth Peer: 'Projekt 2: Notification Management (SimpleSurfaceSocialMedia)'. Student project at course Current Topics in Human-Computer Interaction. Leibniz Universität Hannover, 2022

Contents

CHAPTER 1

Introduction

"There is more information available at our fingertips during a walk in the woods than in any computer system, yet people find a walk among trees relaxing and computers frustrating." Mark Weiser [Wei91]

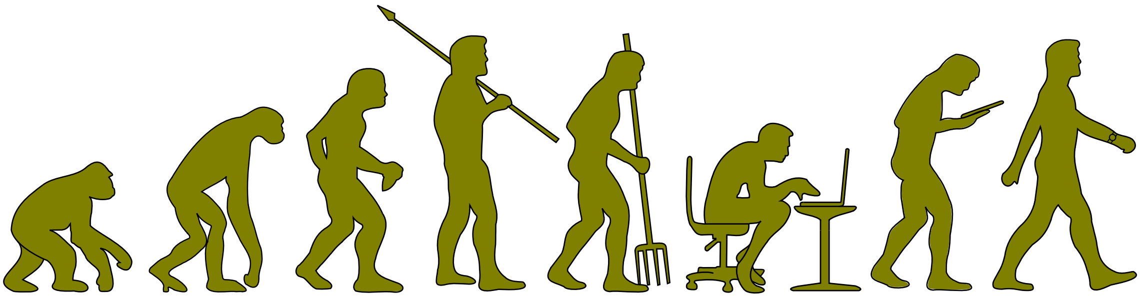

Figure 1.1: An illustration of how technological advances have contributed to human evolution and how the 21st century may contribute to it.

Technology weaves into our daily life. The smartphone has become an indispensable and ubiquitous mobile device [Bal06] enabling worldwide communication and making information available at our fingertips. The term *ubiquitous computing* originates from Mark Weiser who envisioned that computers will *"weave themselves into the fabric of everyday life until they are indistinguishable from it."* [Wei91]. Instead of sitting in front of a PC, computers of various sizes in the environment should assist people in their daily tasks. More than 30 years have passed since Weiser's article, and the number of transistors on integrated circuits has doubled on average in less than every two years since then, known as *Moore's Law* [Moo65; Ros13]. This technological progress has enabled to integrate complex computing systems into everyday objects like watches. Furthermore, virtual and augmented reality headsets can now be used in a mobile context. This means that Weiser's proposed computers of various sizes do not need to be placed in the environment, but can be displayed virtually in the user's field of vision. All that is required now is an ordinary surface on which the information can be displayed. In addition, a surface in itself contains a variety of information that can be used for interactions. Weiser may have been aware of this when he pointed out the wealth of information at our fingertips.

Surfaces surround us everywhere and unconsciously influence our behavior. In shopping centers the operators take advantage of human perception by using different floor materials to persuade customers to make more purchases. For this purpose, the velocity of a pedestrian is influenced by using different soft and hard floor materials [Pre73]. The slipperiness of a floor can be perceived via both visual and tactile cues [Li19], resulting in a changed gait pattern [Cha02]. Consequently, customers can be unconsciously slowed down when walking past the goods. Besides visual and tactile stimuli of a surface, the sound when stepping on a floor can also influence the walking pace [Tur13b]. This shows that we perceive surfaces in our environment mainly through the three classical senses vision, touch and hearing. As a result, surfaces around us can be actively shaped to influence our perception and behavior. Another application field is guidance for visually impaired individuals in urban environments. Pedestrian walkways partially have special cues for the blind. Almost unnoticed by people without visual impairments, tactile pavings are textured paving stones that convey cues for junctions and road crossings [Sak00]. Figure 1.2 shows an example of tactile pavings in the city center of Hildesheim, Germany, at a bus stop. The direction of the grooves in the tactile pavings indicate the walking direction. The curb contains small bumps to delimit the sidewalk and grooves on the road signal the transition to the other side of the road. A person can actively perceive and interpret the unevenness in the stones by using a blind stick. Besides guidance, text can also be displayed via tactile cues in embossed paper or other surfaces. Braille is a tactile font based on dot patterns stamped into the material [Fou13].

Figure 1.2: Tactile pavings in an urban environment near a bus stop. The different grooms in the stone indicate the walking directions. Little bumps on the curb limit the walkway.

In human-computer interaction, surface computing places interfaces on everyday objects like tables. For example, smart tables in home environments [Kir12], in public places such as museums [Row14], or in office environments [Rek99] can enhance the shared experience and enjoyment of using technology. In addition, methods were proposed to make virtual textures on the display tangible through electro-tactile effects [Bau10]. Many approaches of surface computing are limited to the display size or an permanently equipped environment, which in turn constrains the ubiquitous use of technology. We can now, for example, virtually project interfaces onto any surface and integrate surface properties into interactions with mobile computers. This makes it possible to transform any environment into an interactive space that can assist people in their daily tasks. However, surface computing has not yet distinguished between permanently installed systems and ad-hoc approaches.

Therefore, this thesis further subdivides surface computing into *active surfaces* and *passive surfaces*. An *active surface* is based on embedded equipment in or around a surface. The term *active* refers to intentional changes made to a surface to enable interactions between humans and computers. Active surface interactions include, but are not limited to, QR codes, NFC tags, displays (embedded screens or permanently mounted projectors), and embedded hardware to recognize touch events. In contrast, a *passive surface* utilizes and augments the features of an ordinary surface for interactions with mobile technology. The term *passive* refers to an ordinary surface that has not been modified in any way to make it usable for interactions with computers. Passive surface interactions include, but are not limited to, projected interfaces and information via AR headsets, touch events detected via mobile hardware or augmentation and measurement of surface characteristics to convey and retrieve information.

1.1 Hypotheses

This thesis investigates how passive surfaces can be integrated into human-computer interactions to support users of mobile devices in their daily tasks. The aim is to turn any environment and it's surfaces into an interactive space without equipping the environment with technology. For this purpose, various hypotheses are to be formulated and answered. As a first step, previous research in the area of surface computing must be analyzed for a potential subdivision into active and passive surface computing. Thereby, we analyze the following hypotheses:

H1 Surface computing can be subdivided into active and passive surface computing.

H2 Active surface interactions are convertible into passive surface interactions and vice versa.

As already mentioned, we perceive surfaces mainly with our three classical senses: vision, touch and hearing. Sensors for mobile computing systems can also measure these stimuli of the environment and surfaces. To name a few examples: Camera systems can capture textures and colors of a surface [Yeo17], inertial measurement units can sense vibrations related to the roughness of a surface [Hsu15], and microphones can sense sounds related to friction on a surface [Har08b]. From the use of these sensors for passive surface interactions, the following hypotheses arise regarding the user performing an interaction:

H3 Passive surfaces contain information about the type of interaction that is performed on it.

H4 Passive surfaces contain information about the user that performs an interaction.

To investigate these hypotheses, we particularly focus on vibrations and scratching sounds produced by the friction on a surface. We examine these surface related features individually as well as in combination with each other. We aim with the following hypothesis to find out whether surface properties contain complementary information about an ordinary surface and the interaction performed on it.

H5 Vibrations combined with scratching sounds produced by friction on a surface contain complementary features that contribute to identify an interaction.

H6 Interactions with passive surfaces emit information about the type of surface on which the interaction is performed.

Based on the hypotheses mentioned above, we look at specific application cases in which surfaces can possibly support people in everyday tasks. From the presented applications, the following hypotheses will be investigated.

H7 Passive surfaces and their properties in the wild can support users in their everyday tasks.

H8 Passive surfaces are able to sense the context of using technology.

Lastly, this dissertation will look at the augmentation of ordinary surfaces in the wild. The aim is to implement an interactive space anywhere in which we further investigate the role of passive surface computing. We here focus on acoustic augmentation in urban areas and observe the influences on the human perception and behavior. From this, two last hypotheses can be formulated.

H9 Acoustically augmented surfaces influence the perception of an overall perceived reality.

H10 Acoustically augmented surfaces can communicate special cues and change the behavior of the user.

In order to accept or reject the formulated hypotheses we use commonly applied research methods in human-computer interaction [Mac12]. As a research vehicle, we present novel prototypes that we have built and evaluated in various laboratory and field studies. In addition to quantitative data, we collected qualitative data in questionnaires and interviews before, during and after the studies were conducted.

1.2 Contributions

This thesis introduces a sub-categorization of surface computing into active and passive surface computing. Based on this definition, prototypes and interaction techniques are presented to integrate passive, ordinary surfaces into human-computer interactions. Furthermore, scratching sounds and vibrations produced by friction when moving over a surface are investigated in detail in various application scenarios. The three main contributions are listed below.

(1) Definition of active & passive surface computing

(2) An in-depth investigation of scratching sounds and vibrations on passive surfaces in various application scenarios

(3) Methods for integrating passive, ordinary surfaces into human-computer interactions

1.3 Thesis Outline & Prototypes

To investigate the presented hypothesis, various prototypes were built and evaluated in user studies. The following chapters were sorted based on the involved hypotheses and prototypes. We now give an outline of the next chapters and how they contribute to the hypotheses.

Chapter 2 first deals with human perception and the classical senses of vision, hearing and touch. We give an insight into how environmental stimuli are transduced into electrical signals and where these signals are processed in the brain. Furthermore, it is shown how the respective senses contribute to the perception of surfaces. We then continue with a categorization of approaches from previous research in the field of HCI and surface computing into active and passive surface interactions. This serves to test **H1** and **H2**. In a summary we point out limitations of active and passive surfaces.

In Chapter 3 we present a pen prototype to measure vibrations and scratching sounds during handwriting as the pen tip strokes on paper. Using the application scenario of handwriting recognition on ordinary paper, we validate **H3** and **H5**. Furthermore, an approach is presented to efficiently combine the benefits of audio and motion data for handwriting recognition.

We continue the application scenario of handwriting recognition in Chapter 4. Here we mainly test **H4** by identifying writers from individual handwritten symbols.

In addition, we again verify **H3** using an iteration of the pen prototype from Chapter 3.

In Chapter 5 we conclude our pen investigations by evaluating magnetic tracking systems for stylus pens. Here we focus on around device interactions for tiny wrist-worn displays. The aim is to include the surrounding skin into the interface for avoiding display occlusions. We present a pen prototype with built-in magnet and touch detector, which is localized by the embedded magnetometer of a smartwatch. In a small usability experiment we evaluate the back of the hand as a suitable drawing space and compare it to 3D mid-air interaction techniques.

Chapter 6 is dedicated to surface vibrations and scratching sounds in the wild. We investigate **H5**, **H6** and **H7** in two application scenarios. For this purpose, we present a universal sensor unit that can be attached to various devices. In a study we evaluate whether vibrations and scratching sounds can be used to measure different road surface conditions in a cycling scenario. We then evaluate our sensor unit in a skiing experiment where the sensor unit is attached to a ski boot.

We continue to evaluate **H7** in Chapter 7 by using visual features of surfaces. The goal is to measure the context of smartphone usage by identifying different surfaces in a home environment. A prototype is presented and evaluated that measures various household surfaces using rear cameras and flashlights. We further envision various applications that serve to test **H8**.

In chapter 8, an augmented reality application is presented to facilitate the search for specific books in bookshelves. The mobile application uses color and text features of book spines to identify a desired book and highlights it in the camera image view. We further investigate **H7** by comparing search times for books on the shelves with our application against normal visual search. Furthermore, we envision an application scenario for public bookcases to link online communities to local sharing places.

Chapter 9 deals with augmented surfaces in urban environments. An application is presented that acoustically augments footstep sounds during walks in city centers. The listener wears headphones with active noise cancellation while being exposed to various soundscapes and footstep sounds. An in-situ field study serves to evaluate different effects and the individual perception of the listener in a real application scenario. From interviews and quantitative data **H9** and **H10** are tested.

In Chapter 10 we conclude our findings by discussing our results and hypothesis. In a summary, we define limitations of passive surface interactions and outline future research.

CHAPTER 2

Background

The purpose of this chapter is to work out the differences between interactions on active and passive surfaces. We will first briefly look into cognitive psychology to give a basic understanding of human perception. This is followed by a description of how our senses transduce and perceive surface-related environmental stimuli. Then we will continue by defining active and passive surfaces and how they are used in human-computer interaction. To summarize, this chapter will end by clarifying the differences between active and passive surfaces.

2.1 Human Element

At every moment when we interact with the environment and computational systems, several processes take place in our brain. Therefore, the human element in interactions is an important aspect that needs to be addressed for this thesis. In the following, we will give some fundamentals of cognitive psychology and interactions with the environment. To understand how we perceive environmental stimuli, we imagine a simple birthday scenario in Figure 2.1. On the birthday table is a birthday cake with a candle, a gift, a flower and a radio playing music. Several processes now take place in the brain in a very short time that give the person a picture of the overall reality. The intended goal is to blow out the candle for making a wish, cut and eat the cake, as well as open the birthday gift.

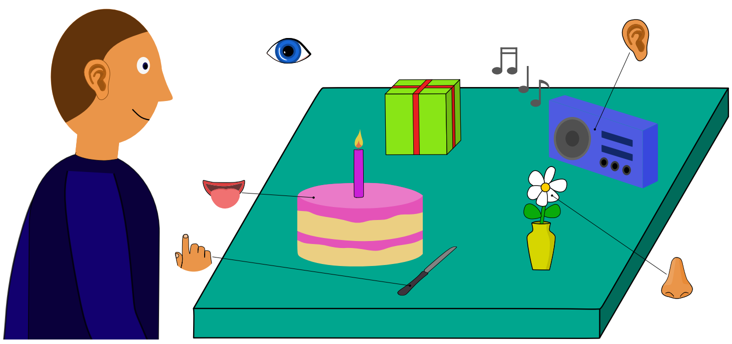

Figure 2.1: Birthday scenario: A visualization of different environmental stimuli and the related classical senses.

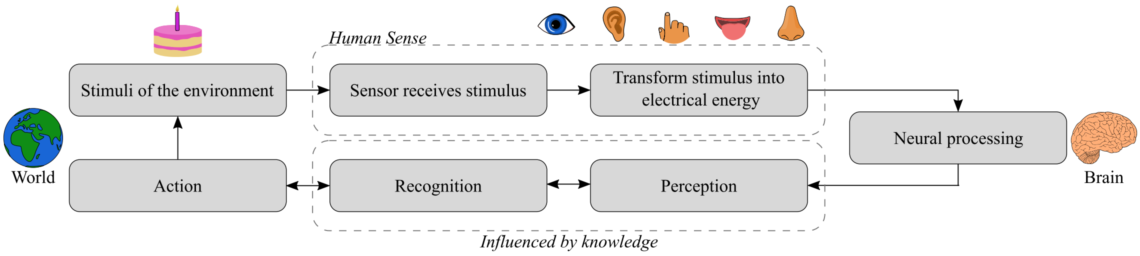

Figure 2.2: Simplified perceptual process adapted from [Gol17].

2.1.1 Perceptual Process

In general, the perceptual process in Figure 2.2 is a continuous cycle that starts from an environmental stimulus and leads to interaction with the environment. We here briefly describe the process adapted from [Gol17]. To exemplify, we take the visual stimuli of the birthday cake in Figure 2.1 and describe the process of perception shown in Figure 2.2.

In the first step, light is emitted by the candle and reflected by the cake into the environment. When the light reaches our visual sensor, e.g. the eyes, light is transformed into electrical energy by the corresponding sensory receptors. Regardless of which sense responses to an environmental stimulus, the transfer from one form of energy to electrical energy is called transduction. The resulting electrical energy is processed towards the cerebral cortex which is a 2 mm layer on the surface of the brain. Figure 2.3 visualizes in which part of the brain each sense is mainly processed. For vision, the most processing occurs at the occipital lobe. In this step the signals of all receptors travel though a network of interconnected neurons where they are attenuated or amplified. The result of that processing step is the perception of the object and the final recognition of the birthday cake with the candle. To further distinguish: While recognition categorizes the birthday cake, perception is the conscious awareness of the object. The process of perception and recognition is not necessarily a sequential procedure. It can also happen in the reverse order and simultaneously [Gib94].

In the last step the person performs an action in the real world to interact with the perceived object. In our case, an imaginable interaction would be to blow out the candle to make a wish, and then slice the cake and eat a piece. Each interaction with the world results in new stimuli measured by the human sensors that are again further processed. Thus, the perceptional process is continuous and dynamic. Another factor that influences this process is knowledge. In our example, the person knows that it is his or her birthday and recognizes the candle on the cake as a sign for being a birthday cake. Looking at the whole scene, the person might recognize other signs such as the smell of the flower or the song that is played on the radio. These environmental stimuli gain importance in case of visual impairments and have an impact on the perception of a scenery. For instance, visually impaired people prefer different landscapes than people without visual impairments as they focus more on their acoustic perception [Sou69].

2.1.2 Senses

After introducing the perceptual process, we will now look into the senses that transduce environmental stimuli into electrical signals. From a historical point of view, Aristotle defined five senses that are responsible for perceiving our environment [El 11; Gru08]: vision, hearing, touch, smell and taste. Since the classical senses convey stimuli of the environment, they are called exteroreceptors [Feh17]. In human-computer interaction, vision, hearing, and touch are the main senses involved [Dix04]. This also applies to surfaces such as touchscreens. We see the displayed information and know from the haptic feedback [Bic08; El 11] and clicking sounds [Mac12] when the screen is touched that an interaction was successfully performed. Penfield was the first to distinguish motor and sensory areas in the cerebral cortex [Pen50]. Higher level of senses, such as satisfaction or frustration, for example reflect how people experience their interactions with computers [Mac12]. In the following, we describe how the senses, that are mostly related to the perception of surfaces, transduce the environmental stimuli. A deeper insight is given to make the perception of surfaces in the environment comprehensible. The descriptions are mainly based on [Feh17; Gol09; Gol17]. Figure 2.3 provides an overview of where most of the processing for each of the five senses happens in the brain.

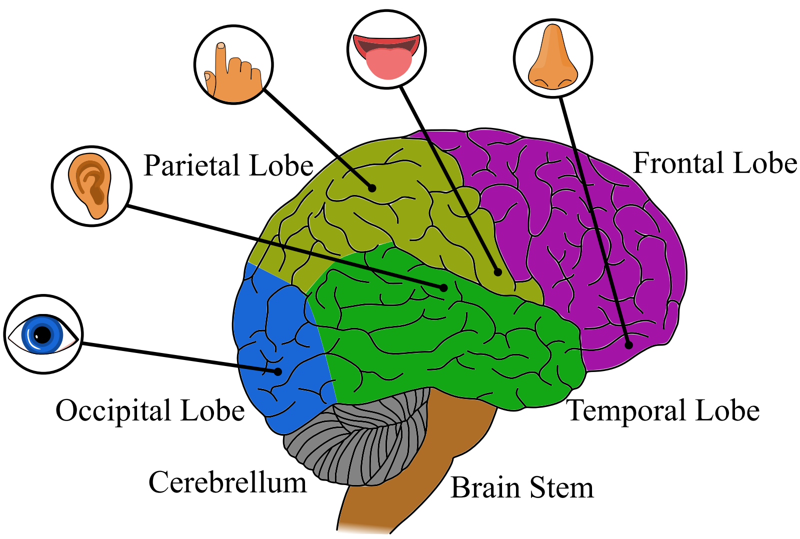

Figure 2.3: A pictorial drawing of the brain areas that are mainly involved in processing each of the senses. The self created drawing is based on [Gol17].

2.1.2.1 Vision

Visible light are electromagnetic waves with a wavelength between 400 nm to 700 nm. The eyes are responsible for sensing light from the environment and transducing the incoming energy into electrical signals. In addition to the general visual impressions, the slightly different perspectives of the eyes enable the three-dimensional perception of the environment with both eyes. Figure 2.4 on the left depicts a horizontal cut of the eye. The dioptric apparatus are the components of the eye that are responsible for refracting incident light [Sch11]. First, the light travels through the *cornea*, then the *anterior chamber* on to the *iris*. The *iris* operates similar to a camera aperture and regulates the incidence of light on the *pupil*. To avoid overexposure or underexposure, the *iris* changes its circular opening width to the *pupil*. The *pupil* appears like a black hole when viewed from the outside, since the *retina* completely absorbs the incident light. Behind the pupil, light falls on the *lens*, which adjusts its shape via muscle fibers (*zonules*) for focusing the light to the distance of the object being observed. The focused light beam travels through the *vitreous chamber* to the point of sharp vision on the *retina* (*Fovea centralis*).

The general structure of the *retina* is shown in Figure 2.4 on the right. The main function is to transduce the light energy into electrical energy. The *retinal pigment epithelium* (RPE) is attached to the choroid and adsorbs incoming light to attenuate noise of the optical system and improve the image quality [Str05]. Above the RPE are the photoreceptor cells which are responsible for the visual phototransduction. *Rod cells* are very sensitive to incoming light, have a slow response time and handle the scotopic vision. The human *retina* includes on average 92 million *rod cells* which can be mostly found at the retinal edges [Cur90]. Therefore, *rod cells* are mainly responsible for peripheral vision and play a subordinate role in color vision. *Cone cells* are less sensitive to light and are responsible for color vision. Hence we can hardly distinguish colors at night. In total, humans have on average 4.6 million *cones* [Cur90] which have a fast response time and are concentrated around the *fovea*. There are three types to cover the visible spectrum. The S type *cones* have blue receptors (sensitivity maxima at 420 nm), M type *cones* cover the green spectrum (534 nm) and L type *cones* are sensitive to red light (564 nm). Photoreceptor cells constantly release glutamate. When light falls on photoreceptor cells, they release less glutamate which is received by the *bipolar cells* that transmit sustained graded potentials to the *ganglion cells*. *Bipolar cells* can ether forward or invert light energy signals. The retina consists of about 36 million *bipolar cells* [Feh17]. The less *rods* and *cones* are connected to a *bipolar cell*, the higher the image resolution. *Horizontal cells* are interconnected between *rods* and *cones*. They serve to amplify light signals. *Ganglion cells* collect the output signals of the *bipolar cells* and form action potentials that travel though the *optic nerve* to the brain. The point where all *ganglion cell* axons leaving the eye is called *optic disc*. There are no *rods* or *cones* that recognize visual stimuli. Thus, this point is also called blind spot. In the brain, most visual stimuli are processed at the occipital lobe.

Although research has uncovered a large part of the processes involved, it is still unclear how exactly the brain perceives visual stimuli [Feh17]. It remains of research interest how we exactly visually perceive and identify surfaces [Ade01]. However, surfaces can subconsciously influence our behavior. For example, a hard, glossy walked surface is perceived as slippery and causes people to unconsciously walk more slowly [Pre73]. In general, gloss indicates the reflectivity of incident light on a surface. The reflectivity is related to the proportion of reflected light (albedo). The perceived glossiness of a digital black and white image was associated with the luminance histogram distribution of the pixels [Mot07]. It was suspected that the human brain adapts image statistics to perceive gloss of surfaces [Mot07]. However, this hypothesis was rejected since the reflected images on the surface and other stimuli of the environment also contribute to the gloss perception [And09; Lan07]. There are six types of perceptual gloss: surface texture (smoothness of a surface), sheen (gloss at grazing angles), specularity (shininess), haze (milky effects), contrast (contrasts between specularities), and distinctness (sharpness of reflected image) [Cha15; ODo86]. In addition, studies show that movement (optical flow) influence the perception of a surface material and overrides static cues [Doe11; Mao19; Wen10].

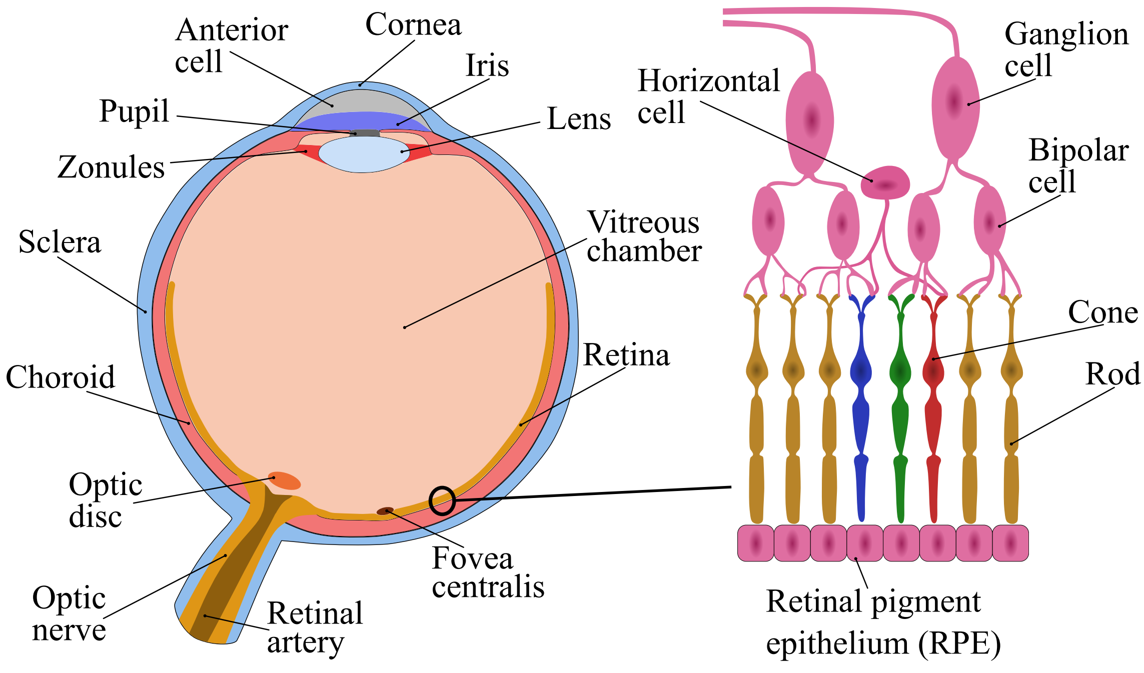

Figure 2.4: A pictorial drawing of the human eye on the left and a sketch of the retina on the right. The drawings are self made based on [Feh17; Gol17; Wik13].

2.1.2.2 Hearing

Acoustic waves are vibrations that propagate through a transmission medium. Our ears are responsible for sensing acoustic waves. Humans are most sensitive to frequencies between 500 and 5,000 Hz but can in the best case recognize tones from 10 up to 20,000 Hz [Feh17]. A tone is a single, periodically repeating acoustic wave. The pitch indicates how high a tone is perceived. Sounds are a composition of different waves with different volumes, which are perceived simultaneously. We can localize sound sources by the delay of sound waves reaching the ears and the attenuating sound shadow of the head.

Figure 2.5 depicts a simplified drawing of the ear. Acoustic waves reach the *pinna* and are bundled towards the *external auditory canal*. The outer ear ends at the *tympanic membrane*. The task of the middle ear is to transform the air pressure of the acoustic waves into fluid pressure waves. Air pressure waves cause the *tympanic membrane* to vibrate. The vibrations are transmitted towards the three ossicles in the *tympanic cavity* that serve as a form of impedance matching from air pressure to fluid pressure waves. Attached to the *tympanic membrane* (eardrum) is the *malleus* (hammer), that transfers it's movements to the *incus* (anvil). The *stapes* (stirrup) are connected to *cochlea* by an oval window and are the final step in the impedance matching process. The *eustachian tube* connects the *tympanic cavity* with the oral cavity and ensures pressure equalization on both sides of the *tympanic membrane*.

The inner ear starts at the *cochlea* (Fig. 2.5, middle) which is filled with fluid and holds the actual sensing *organ of corti* (Fig. 2.5, right). Its task is to transduce fluid pressure waves into nerve impulses. The *semicircular canals* recognize rotations of the head and are thus responsible for the sense of balance, linear acceleration, rotation and gravity. Its electrical signals are transmitted via the *vestibular nerve* to the brain. The *cochlea* is constructed as a spiral around the cone shaped bone *modiolus* where the fluid waves are propagating in different chambers. The three chambers are divided by the *basilar membrane* and the *Reissner's membrane*. The two outer chambers *scala tympani* and *scala vestibuli* are filled with *perilymph*. The inner chamber *scala media* is filled with *endolymph*. When the foot of the stapes stimulates propagating pressure fluid waves, they travel along the *scala tympani* up to the *helicotrema* which is connected to the *scala vestibuli*. At the far end, the waves reach the *round window*, which serves to compensate the vibrations to avoid permanent fluid wave loopbacks along the compartments. The auditory transduction of the waves into electrical signals happens in the *organ of corti*. The fluid waves stimulate the *basilar membrane* to vibrate which causes the hair cells to be deformed when touching the *tectorial membrane*. This causes an chemical process with the *endolymph* that generates electrical impulses which are transmitted via the *auditory nerve*. There are two types of hair cells. The *outer hair cells (OHC)* are shaped like rods and act similar to an active amplifier of the acoustic signals. They can change their dimensions according to the membrane potential with a frequency of up to 24 kHz [Feh17]. The *inner hair cells (IHC)* transduce the movements along the *tectorial membrane* into electrical impulses. These electrical impulses are transmitted by the *spiral ganglions* via the *cochlea nerve* to the brain's temporal lobe. *IHCs* are frequency selective because the *basilar membrane* becomes progressively larger along the cochlea. This causes frequency resonance to decrease from the base of the *cochlea* towards the *helicotrema*.

The structure of the cochlea thus allows multiple external auditory stimuli to be sensed simultaneously. The highest sensitivity can be found in the range of speech [Feh17] which is also an important interaction modality [Mac12]. Our brains are capable of filtering noise sources while focusing on specific sounds, which is known as selective directional hearing or the *"cocktail party effect"* [Che53]. We can walk through a forest and focus on different sound sources. Walking meditation is a form of concentration on one's gait that takes advantage of this phenomenon [Han85]. The practitioner is encouraged to focus all senses on the gait, which includes the sound of each footstep. Research has shown that footstep sounds can contribute to a person's presence in virtual worlds [Han85; Kab82; Kab13; Ker20; Ser11]. Furthermore, augmented footstep sounds were found to influence walking behavior [Hop19; Ley19; Men10; Ser11; Taj15; Tur15; Tur13b; Wil20] and serve as an unobtrusive feedback channel [Gom20; Ohn19; Rod14]. Besides augmenting surface-related sounds, there has been research on stimulating the human sense of balance by galvanic vestibular stimulation (GVS) [Fit99; Mae05]. By stimulating the organ with electrical impulses behind the ear, the brain can not correctly perceive the position of the head. As a result, the user leans toward the stimulated electrode and cannot maintain straight forward motion.

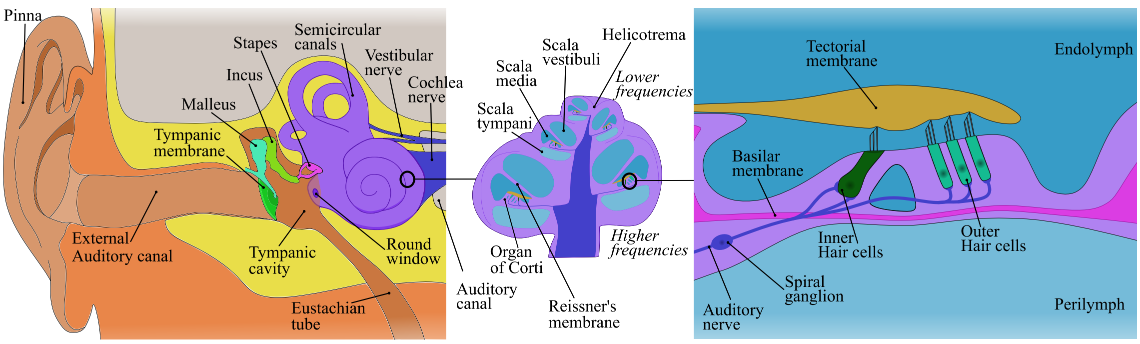

Figure 2.5: A pictorial drawing of the human ear. On the left, the anatomy of the outer, middle and inner ear. The middle picture shows a horizontal cut through the cochlea and the right drawing shows the organ of corti. The drawings are self made based on [Feh17; Gol17; Wik09; Wik04].

2.1.2.3 Touch

Touch refers to a haptic perception that is conveyed by sensors within the skin. Cutaneous or tactile receptors in the skin sense different environmental stimuli such as temperature, pain or pressure [Gol09]. The skin is our largest organ and responsible for multiple functions. Besides sensing external stimuli, the skin ensures temperature regulation and protects our body from external hazards such as bacteria. We will here focus on external stimuli and their perception. In the brain, the sensations of the skin are mostly processed in the parietal lobe [Cla94].

A simplified sketch of the skin is shown in Figure 2.6. We left out several components such as arteries, veins, melanocytes and sweat pores to focus on haptic perception. The sketches are based on [Gol09; Zim14]. The skin can be divided into three main layers: *epidermis*, *dermis* and *hypodermis*. The *epidermis* is the outer layer of the skin that protects the body from environmental hazards like bacteria or chemicals and is responsible for the color of our skin. It is mainly composed of cells that produce keratin to maintain the protective function [Bar12b]. These cells are called *keratinocytes* and migrate up through the *epidermis* until they form an outer waterproof, protective layer named *stratum corneum* [Jan17]. The *stratum corneum* consists of dead cell layers and its thickness differs from an average of 6 layers on genital skin to 86 layers on the heels depending on the part of the body and its exposure to mechanical stress [Ya-99]. In the *stratum granulosum* below, the *keratinocytes* gradually transform into dead cells named *korneocytes*. The *stratum spinosum* underneath contains *Langerhans cells*, which contribute to the skin's immune system [Cho09]. The *stratum basale* is an unicellular layer of basal *keratinocyte* stem cells that are constantly dividing to renew the outer layers [Jan17]. The *basement membrane* tightly connects the *epidermis* and *dermis* below. The *dermis* has various tasks such as temperature regulation, blood distribution and sensing environmental stimuli. The *hypodermis* is the deepest layer filled with fat cells and connective tissue. It serves the body as a storage of energy and water as well as a protection against cold.

Cutaneous receptors can be divided into mechanoreceptors (pressure), thermoreceptors (temperature) and nociceptors (pain, damage). There are several different mechanoreceptors, which are responsible for sensing different external stimuli. They have in common that they do not require additional sensor cells for firing action potentials, but can be divided into fast and slow adapting types. Figure 2.6 on the right visualized three mechanoreceptors. *Merkel cells* sense vertical pressure and are therefore highly involved in the perception of touched surface textures [Zim14]. They are slowly adapting (SA-1) to stimuli and are located at the *stratum basale*. The second slowly adapting type (SA-2) are *Ruffini's corpuscles* located in the *dermis*. They contribute to touch-pressure sensing as well as perceiving self-movement and body position. *Meissner's corpuscles* are located directly underneath the *stratum* *basale* in the *dermis*. They are rapidly adapting (RA) to stimuli like pressure and low-frequency vibrations. *Pacinian corpuscles* are free nerve endings enclosed in a layered capsule that rapidly adapt (PC) to high-frequency vibrations. *Free nerve endings* are nociceptors that sense itch, pain as well as hot and cold temperatures with a slow response. Nerve endings around the hair follicle are fast adapting to external stimuli and recognize touch as well as motion of the hair.

Tactile perception of surfaces can be divided into five dimensions: macro roughness, fine roughness, warmness, hardness and friction (stickiness) [Oka12]. Touch is the act of bringing a part of the body into contact with an physical object. In general, touch can be divided into two modes: First, active touch refers to an intended environmental exploration of an individual. Second, passive touch refers to the act of being touched by something or somebody [Cha09]. Both modes can be further subdivided into static (no movement along a surface involved) and dynamic touch. Dynamic touch is twice as sensitive in perceiving the roughness of a surface [Mor83] since static touch involves less cutaneous mechanoreceptors (SA-1, SA-2) [Cha09]. A static touch event can for example convey information about temperature, surface texture and the shape of an object. Active touch events like contour following and texture exploration usually require relatively slow movements with low forces of about 0.5 N [Smi02]. It has been argued that perceptual performance is lower with passive touch because it refers to a more unnatural experience [Gib62]. Active touch has been shown to be more efficient in texture exploration. However, with increasing exploration time, passive touch achieves similar performance [Cha94]. When exploring a surface, the inter-sensory interaction of vision and touch is important for the perception [Led09]. For spatially defined textures, vision plays a greater role, while the sense of touch becomes more important as surface roughness increases [Led86].

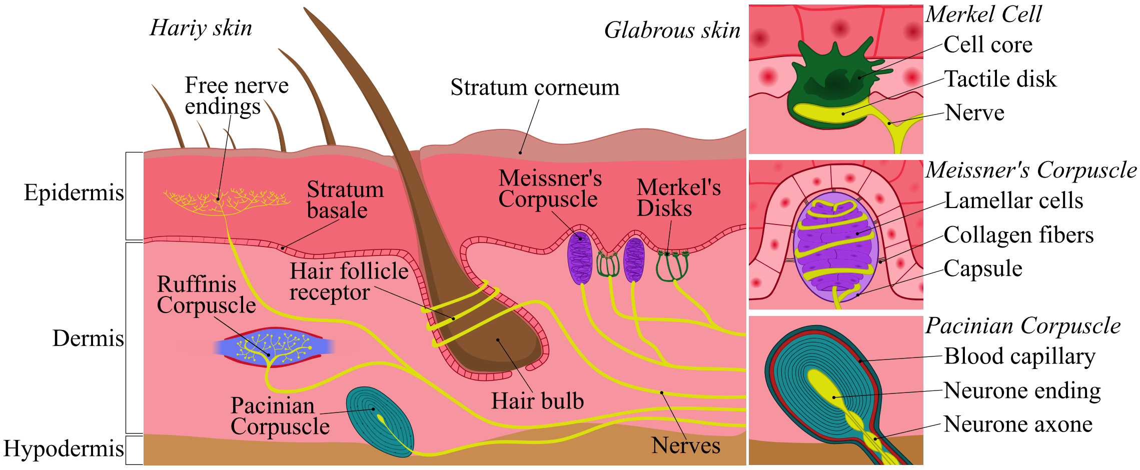

Figure 2.6: A drawing of the human skin on the left and a sketch of the touch receptors the right. The drawings are self made based on [Feh17; Gol09; Wik17; Zim14].

2.2 Ubiquitous Surfaces

Surfaces surround us all the time and everywhere, or in other words: surfaces are ubiquitous. As we have already discussed, their properties contribute to the perception of the environment. In human-computer interaction, surfaces are used for various input and output modalities [Mac12]. For example, a touchscreen can be used to display information and recognize (touch) input events. Innovations in hardware miniaturization, wireless communication and decreasing power consumption have enabled to embed technology in or on all kinds of materials. Mark Weiser and his colleagues at Xerox PARC were the first to envision computers *"fade into the background"* [Wei91]. Instead of personal computers, smart everyday objects equipped with interconnected computers embedded into the real world will unnoticeably support humans in their daily tasks. Weiser exemplified that a room is filled with many different writing and display surfaces that can lead to hundreds of invisible computers.

His article inspired other companies and researchers to establish competing alternative research agendas of ubiquitous computing [Dou11]. For example, *proactive computing* (Intel) is a concept of sensor and actuator networks in the environment that *"monitor and shape their physical environment"* [Ten00]. Here, the focus is shifted towards artificial intelligence, data processing and user experience. *Ambient Intelligence* (Philips) aims to embed sensors and actuators in the entire environment to individualize and adapt computational systems to each user [Aar09; Aar03; Den01]. *Pervasive computing* (IBM) focuses on mobile devices and business processes [Ark99; Mat01]. Abowd and Mynatt introduced *everyday computing*, which *"promotes informal and unstructured activities"* that are *"continuous in time, a constant ebb and flow of action that has no clear starting or ending point"* [Abo00]. They motivated that interfaces should be designed more naturally and use a greater variety of communication capabilities between humans and computers. In addition, ubicomp applications should consider the context of their use based on sensed information from the real and digital environment. A system can be described as context-aware if the information and services provided are adapted to the current user task [Abo99]. One example are navigation tasks, where the displayed route is adapted based on road information [Mat13]. In Chapter 6, we will show factors of road surface conditions that influence preferred route choices. In addition, we will present a vision-based approach to identify surfaces in Chapter 7. Here, identified surfaces will serve to recognize the context of smartphone usage.

In general, this thesis distinguishes between two types of surfaces that can be used in a ubiquitous interaction: Active surfaces, in which technology is directly embedded into the surface or permanently installed in the vicinity, and passive surfaces, which use mobile technology to measure and integrate surfaces into interactions. To distinguish from other fields of research, an active surface in medical science, for example, refers to a chemically or biologically reactive implant surface that reacts with the physiological environment [Bos12]. Large radio telescopes correct imperfections by adjusting their shape which is called active surface technology [Wil21]. More generally, while passive surfaces have static properties, active surfaces can dynamically adapt their properties to their environment and a specific use case. We will now further describe the usage of active and passive surfaces in HCI.

2.3 Active Surfaces

Active surfaces have played an important role in the development of intuitive user interfaces decades before Mark Weiser's article about ubiquitous computing [Wei91]. A screen can generally be considered as an active surface that changes its properties (appearance, color). Early computers used buttons and keyboards to make inputs while lamps or screens displayed the current state of the system [Mac12]. The first computer that allowed interacting directly on the screen was *Sketchpad* introduced by Ivan Sutherland in 1962 [Sut64]. The computer used a light pen to draw, resize, move or delete objects directly on the screen. 15 years later, Alan Kay and Adele Goldberg presented *Dynabook* which was a portable computer that allowed stylus input [Kay77]. In both systems users were able to intuitively interact directly on the screen with displayed objects similar to manipulating physical objects. This is known as *direct manipulation* and was introduced by Ben Shneiderman [Shn81; Shn82]. In the same time period, the first touch displays for direct finger input were invented [Joh65]. The *Xerox Star* computer made another breakthrough in 1981 by integrating the concept of an office desk into the digital interface known as the *desktop metaphor* [Mac12]. One of the first systems that aimed to directly connect the virtual desktop with the physical desk was *DigitalDesk* by Pierre Wellner

[Wel91] in 1991. Users could interact with digital and physical paper on the desk by a front-projection including finger and pen tracking using a camera system. This milestone marked the beginning of tabletop interaction and surface augmentation, which we will briefly outline in the following.

2.3.1 Display Technologies

The display is responsible for presenting the interface and system state on a surface. A differentiation can be made between projected and embedded displays [Sch10]. Projected displays on active surfaces refer to a fixed installation near the surface. A full review of display technologies is given by Walker [Wal12]. Digital projectors can be ether used in a rear or front projected setting. Front projected settings can embed interfaces on a variety of surfaces, but will cause occlusion if objects or hands come into the light beam. The advantage is that front projectors can illuminate uneven and opaque surfaces. Front projections can, for example, augment a working environment and display additional information about objects [Rek99]. In combination with depth tracking, shapes of the surface can be dynamically integrated into the view [Woo16]. Furthermore, depth cameras can serve to distinguish user input [Elm12]. Projections from the back can make the displayed image on the surface more robust to ambient light. In addition to flat surfaces, back projections are also suited for creating spherical displays [Ben08]. However, in both projection setups a certain distance is required between the projector and surface which is known as *throw*. Furthermore, the *brightness* is often insufficient to use this technology outdoors. Since rear projectors reflect a portion of light from the rear projection surface back to the projector, they usually achieve lower brightness than front projections. Displays embedded directly into the surface usually have better image quality and are partially suitable for outdoor installations [Mot08]. Compared to projectors, they achieve smaller display sizes and, depending on the technology, can be installed as planar or slightly curved surfaces.

2.3.2 Touch Technologies

A differentiation can be made between two types of touch sensing approaches on active surfaces. On the one hand, there are touch technologies that are embedded directly into the surface and on the other hand touch events can be tracked by external permanently mounted hardware. We here briefly describe several approaches. A detailed description is given by Schöning et al. [Sch08a; Sch10]. In addition, Paradiso et al. have summarized some of the early works to create interactive surfaces [Par00].

2.3.2.1 Resistive touch surfaces

Resistive touch surfaces are sensitive to pressure from the input device or finger. They consist of two stacked conductive layers that change their resistance to each other under point pressure. A small spacer between both layers prevents permanent electrical contact. The position of a touch event is determined by measuring the current vertically and horizontally between both layers. The advantages of this technology are low cost manufacturing and high robustness at the expense of lower tracking accuracy [Dow05]. Similarly, interactions on a surface can be detected via weight measurements on the edges of a display. Load sensing can, for example, detect point and click interactions or objects on a table [Sch02].

2.3.2.2 Capacitive touch surfaces

Capacitive touch technologies can be divided into surface and projected capacitive touch approaches. Capacitive surface technology uses four electrodes at the edges of the contact surface to create a uniform electric field. Fingers affect the electric field and consequently the flowing current. However, multi-touch input cannot be sensed with this setup. Projected capacitive touch is an established sensing approach on smartphone displays. Two stacked electrode layers in X and Y direction accurately detect touch events with multi-touch support. The finger affects the electronic field, which in turn changes the capacitance between a local pair of electrodes. The first capacitive touch surface that could distinguish users in collaborative work was *DiamondTouch* in 2001 by Dietz and Leigh [Die01]. Here, antennas in the surface emit high-frequency signals that propagate through the user's body when touched. Receivers mounted on each chair detect which user touches the surface. One year later Rekimoto presented a mesh of in the surface embedded antennas and receivers called *SmartSkin* [Rek02]. The system was able to recognize gestures such as zooming by analyzing the relative positions between detected touch points. In addition, tangible objects on the table could be integrated into the interface when touched.

2.3.2.3 Surface Acoustic Wave Touch Surfaces (SAW)

Surface Acoustic Wave Touch Surfaces (SAW) use ultrasonic waves to recognize touch events. Piezoelectric transducers in x and y direction send acoustic waves on the edges of a glass panel. Reflectors direct the waves across the panel to the piezoelectric receivers. When a soft material like a fingertip touches the surface, the acoustic waves are absorbed. This technology is capable of detecting single and dual-touch events [Sch10]. In addition to SAW, touching a surface generates sounds that propagate through the material. Contact microphones such as piezoelectric transducers attached to the edges of a surface can detect the position of an input event by analyzing the sound propagation delay. This approach can be used to equip walls with touch and gesture input. [Har08b]. Another dimension of acoustic touch is the produced sound (tap and friction) on a surface. It has been shown that produced sounds can distinguish interactions with, for example, fingers or knuckles [Lop11].

2.3.2.4 Optical Touch Surfaces

Optical Touch Surfaces use light and camera systems to recognize input events. Besides approaches that utilize depth cameras [Wil10] or artificial intelligence (AI) to directly detect hands in the camera image [Mur12b; Wig07] two approaches can be distinguished. The approach of frustrated total internal reflection (FTIR) transmits infrared light through an acrylic surface [Han05]. In order to achieve a total light reflection within a surface the refractive index of the inner material must be higher than the index of the outer material and the angle of incidence of the light at the edges must be very small [Sch08a; Sch10]. When a user touches the surface, the refractive index increases and light is reflected toward a camera mounted below the surface. The amount of reflected light is depended on the pressure on the surface [Dav08].

Like FTIR, Diffuse illumination (DI) relies on the analysis of altered contrast values in the image. A transparent plate is illuminated by infrared LEDs and a diffusor on the plate uniformly disperses the light over the surface. Objects and hands near the surface then reflect the light and are detected by the camera. There are also approaches which illuminate the front in order to create interactive floors [Grø07]. In addition, to achieve a more reliable tracking and to sense pressure DI can be combined with FTIR [Aug10].

2.3.2.5 Pen Input

Pen interactions can be divided into two types. The first are stylus pens, which are used on a display for pointing and drawing. The second type are digital pens, which emit ink on the paper and digitize the information simultaneously. While stylus pens use the tracking system of the screen, there are different ways to detect the position of a digital pen on the paper and to digitize the information.

Simple stylus pens can be used on resistive and capacitive touch surfaces. The stylus transfers the user's pressure or electric field to the surface. Active stylus pens include additional hardware to increase the tracking accuracy. For instance active capacitive stylus pens sense the electric field of the capacitive touch screen. They transmit a response to the incoming signals in order to affect the electric field sufficiently to be registered like a finger. This technology can also be used on fingertips to avoid the problem of finger occlusion on tiny displays [Xia15]. Inductive stylus pens use dedicated tracking hardware underneath the display surface. A coil in the stylus induces an electromagnetic field which is detected by a sensor embedded into the display. This technology has the advantage that hover interactions near the surface can also be detected. Furthermore, inductive sensing can also be used on metal surfaces. The pressure caused by the finger bends the metal plate, which can be registered by a coil under the surface [Kas16]. Magnets are a promising approach to expand the interaction capabilities of styluses. This enables to determine the orientation of the stylus by a magnetic field sensor embedded in a smartphone

[Hwa13b]. Furthermore, interactions on surfaces around the device can be detected. However, this is to be considered as passive surface interaction. Furthermore, the pressure of the pen can be derived from the scratching sound that occurs when a pen is stroked on the display [Hwa12].

Digital pens pursue the vision of bridging the analog and digital worlds. Information that is physically inked on a sheet of paper or other material and simultaneously available digitally. A wide range of application scenarios can be identified ranging from note taking and sketching to annotating documents [Ric17]. There are several principles that are related to active surfaces. By printing unique and almost invisible dot patterns on a sheet of paper, cameras embedded in a pen can link the position of strokes to a position on the paper [AB22]. In addition, transparent dot pattern films can also be attached to screens to enable direct interaction on the computer screen [Hof10]. The company Anoto first announced in 1996 that they create a digital pen based on this technology [Sch08b]. Besides dot patterns, external tracking hardware clipped on the paper aims to sense the position of the pen when stroking. For this purpose, the digital pen continuously transmits signals to the tracking device [Sch08b]. The advantage of this approach is that no special paper is required to digitize handwritten information. For example the company Wacom [Com22] uses this technology for their pen Inkling. Zhang and Harrison present an approach to digitizing handwriting with conventional pens. They laminated a conductive layer onto the back of a sheet of paper in order to determine the writing position by the changed electric field during writing and touch events [Zha18]. In contrast, a passive surface interaction with digital pens refers to digitizing handwritten information without equipping the paper with technology. This means that the pen itself must digitize the information. For instance the company Stabilo presented a pen that supports undergraduate students to learn handwriting [Gmb22]. Furthermore, there are approaches that add cameras to the digital pen in order to recognize handwriting [Uch09] or even try to locate the pen position by extracting individual texture features of a paper [Iwa10]. In the Chapters three to five we will present an own technology to recognize handwriting. We will also present further works related to digital pens.

2.3.3 Surface Effects

As previously described, the perception of a surface is influenced by our senses. Human-computer interaction also takes advantage of tactile effects on surfaces to convey information or make virtual experiences more realistic [Bas20]. The most popular methods to create tactile effects on screens are ultrasonic, vibro-tactile and electrostatic feedback. A comprehensive overview is provided by Basdogan et al. [Bas20]. Furthermore, tactile effects are an important support for the blind. For instance, Braille is an international standard for tactile writing in which text is represented by a dot pattern. This font can be provided either by embossed surfaces or refreshable braille displays. Braille also belongs to active surfaces, as it has to be embossed directly into the surface. In addition, people with visual impairments use tactile pavements for guidance in public spaces [Sak00]. We will now briefly discuss methods for augmenting surfaces with tactile effects.

2.3.3.1 Generating Tactile Effects

Vibrations are oscillations that propagate through the material. The most common used type are vibration motors, which generate vibrations through a rotating eccentric mass. Their rotational frequency can be controlled by the voltage, but they have a start-up delay and require a certain threshold voltage for start-up. Voice coil actuators operate similar to loudspeakers and can produce all types of waveforms with little delay, but are bulky. Linear resonance actuators are a slim voice coil design for mobile devices. Due to their low frequency bandwidth, these actuators are suitable for tactile sensations in the low to mid range action frequencies [Bas20]. Piezoelectric actuators are compact piezoelectric plates that can produce high frequency actuations on surfaces. They are often used in prototyping, but have a low durability, which prevents commercial use in mobile devices. Electroactive polymers change their size or shape under high voltage. They can be applied like a liquid to large displays and remain flexible [Bas20]. Electrostatic actuators produce low forces and are suitable for high frequency haptic effects. They consist of electrode grids which produce friction forces when the finger is moving over the display. The effect of electrostatic actuation goes back to an accidental discovery by Mallinckrodt et al. in 1954 [Mal53]. They reported a rubbery sensation when moving a dry finger over a surface under voltage with a thin insulating layer. In 2010 Bau et al. were the first to use this effect of *electrovibration* on displays to design interfaces [Bau10]. These effects can also be used, for example, to make interaction with a pen on a touch display more similar to the sensation of writing on paper [Wan16a].

2.3.3.2 Tactile Surface Applications

Tactile effects offer a variety of applications on surfaces. They range from target pointing over UI interactions up to supporting visually impaired in daily tasks and artificial rendering of textures. During target acquisition in dragging tasks, ultrasonic friction increases performance [Lev11]. This effect can be reproduced with electrovibration [Zha15]. Button clicks are perceived as more pleasant by tactile feedback if the actuator reacts rapidly and in a short period of time [Par11]. The frequency of the actuator has an influence on how the click is perceived. Low frequencies between 125 Hz and 500 Hz are perceived as dull and crisp [Che11]. Such effects can also be reproduced with piezoelectric actuators [Sad20] and ultrasonic feedback [Gue18]. Ultrasonic friction modulation can be used to display the scrolling amount of a slider [Lev11]. Likewise electrovibration offers potential for sliders or dragging tasks due to the wide frequency range and uniform response over the interaction area [Bau10]. In addition, other applications such as the sensing of textures are possible. This allows for a higher degree of realism and greater pleasure in using technology [Lev11]. For example, textures in digital children's books can be augmented with tactile feedback [Lev16]. Communication via text, images or by touching the screen enables emotions to be conveyed [Mul14]. However, the potential for use in visual impairments is limited as electrovibration can only convey information at the size of the fingertip [Xu11]. Presenting Braille by frequency modulation was found to be insufficient to replace traditional embossed surfaces. Similarly, electrovibration was found to be difficult when displaying sharp contours. However, it can assist the blind in replacing the visual sense [Isr12]. Sensory substitution has also been used on photos. Users can touch the shape of an object or person in an image they are currently viewing [Lim19].

2.4 Passive Surfaces

According to our definition, passive surfaces aim to use surfaces for human-computer interactions, but leave the surface in its original state. This also means that no hardware is to be installed in or near the surface. Instead, the user's mobile devices integrate a desired surface into interactions. Therefore, it enables ubiquitous integration of any surface into human-computer interactions and does not limit the interaction area to a fixed display size. Moreover, computers contain various rare earths in small quantities. A more resource-efficient use of these materials can help to avoid supply shortages [Che19a]. One early technological devices that can be attributed to passive surfaces is the computer mouse. It operates on most flat surfaces and it's interaction space is only limited by the display dimensions. The mouse was originally invented by Douglas Engelbart in 1963 [Mac12]. It changed the way how people interact with computers. Instead of typing commands on a keyboard, the user can move the cursor on graphical icons and execute commands by clicking a button on the mouse. The integration of passive surfaces in interactions aims at a completely seamless connection of the physical and virtual world. Feiner et al. were the first to extended Weiser's vision of ubiquitous computing by small wearable "see-though" and "hear-through" displays in 1993 [Fei93]. Their approach was inspired by the first head-mounted display by Ivan Sutherland in 1968 [Sut68]. In 1993, Feiner et al. did not have the technologies required to augment an office desk like *DigitalDesk* by Wellner [Wel91], but were capable of presenting an augmented maintenance guide for a printer. In the same year Fitzmaurice envisioned that users will carry small devices around to retrieve information from the environment. The display of these small devices then acts as the *"personal display onto information spaces"* [Fit93]. Small handheld devices should be used, for example, to find books on shelves in an office environment by highlighting the position of a desired book. In Chapter 8, we will take a closer look at augmented book search with handheld devices. Nowadays we know

Fitzmaurice's palmtop computers as smartphones, which turned into an universal tool for ubiquitous computing [Bal06]. Their highly integrated sensor technology and computing power have made mobile augmented reality possible, although further research is still required [Cha17c]. We will now discuss how passive surfaces have already been integrated into human-computer interactions.

2.4.1 Visual Passive Surface Augmentation

In the visual augmentation of passive surfaces, virtual information is projected onto the surface in the user's field of vision. This resulting space of real and virtual information is known as *Mixed Reality* (MR) [Mil94]. In Milgram's reality-virtuality continuum, this combination of real and virtual information includes the subclasses of *Augmented Reality* (AR) up to *Augmented Virtuality* (AV). While an AR space contains more real information, an AV space is dominated by virtual information. As a result, physical surfaces can be utilized for presenting virtual information throughout the entire MR space. There are two main technical approaches to visually augment a passive surface: video see-through augmented reality (VST) uses displays in the field of vision to extend reality and optical see-through augmented reality (OST) augment the environmental light entering the eye by projections onto a worn transparent surface in the field of vision. In addition, wearable projectors can directly illuminate physical surfaces and the skin [Gru18; Har10]. OST and projections have the advantage of being able to represent reality at the highest resolution without technical limitations, while VST has the advantage of being able to completely replace all visual stimuli in the environment [Ito21]. A review of display technologies for AR and VR headsets is given by Xiong et al. [Xio21].

According to Itoh et al. [Ito21] there are three main challenges in augmenting real environments: spatial, temporal and visual realism. Temporal realism relates to the update rate of the presented image when moving [Ito21]. Especially in predominantly virtual realities and with latencies above 5 ms [Jer09] this can contribute to motion sickness [Cob99]. Virtual realism refers to how realistic a virtual object or surface is perceived. This is related to how real the colors of the virtual object or surface can be represented and whether the shadows match the overall reality and distance of the object [Ito21]. Spatial realism refers to the visual perspective on an virtual object or surface in the real world. For example, if the perspective distortion of a virtual display does not match on a physical surface, it looses realism and is not perceived as being adhered to the surface. Early works attached markers to the surface for aligning objects and displays [Kat99; Rek99]. Markers and QR codes have the advantage that they offer an affordance to communicate with digital media in the real world. Thus, they can be viewed as "entry points" [Roh05]. However, surfaces with markers can be categorized as active surfaces, since the surface is actively changed. The recent trends are shifting towards extracting natural features of the environment from camera images and adapting the perspective of virtual superimpositions on the basis of these features. There are two principles of vision based tracking: outside-in and inside-out [Rol01]. *Outside-in tracking* uses external, permanently mounted cameras to track the position and orientation of the equipment worn on the head. For this purpose, the headsets are equipped with infrared LEDs which can be easily recognized in the camera image. *Inside-out tracking* places cameras within the mobile device. The algorithm determines the position by estimating the motion of the camera based on feature point movements between two images [Nis04]. The advantage of mobile use without complex calibration results in a loss of accuracy [Mon22]. Motion sensors can be used to improve motion estimation [Esh17].

2.4.2 Sensing Passive Surface Touch

Unlike active surfaces, passive surfaces do not have any technology attached that recognize contact with the finger. Body-worn sensors must register surface contacts to turn any surface into an interactive space. In addition, it may be necessary to detect the location of the touch event. Further, the surface itself is initially unknown to the mobile sensing system. This means that the context may decide whether and what interaction with a computational system is desired by a user. In the wild, surfaces can have dynamic properties that influence the behavior of the user. For instance the user might don't want to touch a dirty or wet outdoor table for interactions with a projected interface on it. An interaction can thus be considerably more complex than on an active surface.

Our skin can be considered as a ubiquitous surface, which is a suitable space for input and output events. Tiny, body-worn displays can use the surrounding skin for input to prevent the display from being covered by the finger during an interaction. Input modalities on the skin can be divided into: touch, grab, scratch, squeeze, press, shear, twist and pull [Wei14]. Furthermore the location of an input event can decide which action is to be performed [Ber19]. Depth sensors [Sri17] or distance sensors [Kni14] can be used to detect an input in the near field of a smartwatch. Moreover, body-worn depth cameras can sense touch events in the environment [Har11]. Distance sensors under a smartwatch can measure the stretch of the nearby skin to detect touch events [Oga15]. The acoustic signals when touching the skin [Har10] or a surface [Gon20] can be measured to recognize touch events. Similarly, acoustic signals sent from a ring on a finger can be received by a microphone array on the smartwatch to track the finger's location [Zha17]. Moreover, magnetic rings [Par19a] or magnets attached to the fingernail [McI19] can be localized by smartwatch sensors. The phalanges of the fingers are suitable as input buttons similar to a numpad. The thumb then serves as the input device [Gol99; Sol18]. Radio-frequency [Zha19] or alternating current signals [Zha16c] that propagate through the body can be utilized to detect touch input events. When both arms touch each other, the signal is changed, which is sensed as an input event. In addition, when a finger touches or strokes a surface, the produced vibrations on the finger can be used to recognize touch events [Shi20]. Our skin can generally be considered both an active and passive surface. As already described, our skin contains sensors that allow us to perceive touch. While this could be considered an natural active surface, from an outside perspective, the skin is considered a passive surface without technology. A smartwatch is an active surface limited to its display size. Integrating the surrounding (passive) skin in the interaction increases the interaction radius and capabilities.

In augmented and virtual realities the trend nowadays shifts towards markerless finger and surface tracking based on natural features from cameras. Earlier works used, for example, color features [Lee07], hand contours [Gru01] and depth cameras [Har11; Wil10] also combined with infrared images [Xia16] to track fingers and touch events. However, early works required to calibrate the environment to a given setup with limited mobile capabilities. With universal touch input, sensor units, e.g. on the finger, can detect touch and motion to circumvent camera calibration challenges [Ngu15]. Nowadays, AR headsets like the Microsoft HoloLens or Google Tango project devices also embed time-of-flight depth cameras for sensing the environment and embedded motion sensors to estimate the camera orientation. To enable fully ad-hoc and un-instrumented touch input Xiao et al. presented and approach without the need for calibrations [Xia18]. They first extract the surface planes from the depth image using the RANSAC algorithm [Fis81]. This allows to segment the hands from the surface using the height profile with edge filtering of the infrared image and depth map. After smoothing the found contours, the fingertip is then selected by extracting the convex shape of the fingers. The recognized position of the fingertip is further smoothed over sequential frames. Touch events are detected by thresholding height measurements.

In addition to interactions with virtual interfaces, one vision of un-instrumented touch input is the virtualization of the keyboard. The first virtual keyboards were invented at the beginning of the 21st century. They equipped the hands with motion sensors, integrated touch functionality into gloves, used projected laser displays or attached markers to the hands [Lee03]. Nowadays, AR headsets can project keyboards in the field of vision for typing on surfaces or in the air [Lu21]. However, it has been shown that projected keyboards on a surface reach significant better typing performances [Dud19].

2.4.3 Augmenting Passive Surface Touch

Previously in the section active surfaces we have already shown that augmenting tactile perception is challenging. The same applies to passive surface interactions. In general, we can distinguish three approaches to augment passive surfaces related to our classical senses: Appearance, haptics and acoustics. Fundamental work by Lesaki et al. has shown that superimposing a surface with texture images can influence the tactile illusion. However, this is limited to materials with a similar hardness or softness [Les08]. The same applies to acoustic feedback in active and passive touch interactions [Kan21]. A tactile illusion is thus influenced not only by the sense of touch, but also by visual and acoustic perception.

In particular, the vision of making virtual worlds more realistic has led to intensive research in the field of tactile feedback and perceiving virtual objects or surfaces. A tactile glove can be considered as an active surface layer on the skin. However, if the glove is used to augment the perception of a surface from the (real or virtual) environment, we speak of a passive surface interaction. For instance a tactile glove can sense objects by hand posture and pressure [Sun19] as well as provide tactile feedback on touched objects and surfaces [Mur12a]. Experiments have shown that gloves with electrotactile and vibrotactile feedback combined with thermoelectric feedback can discriminate virtual textures (smooth/rough and soft/hard) at high accuracy [Kee20]. Smart gloves can also have real-world applications. For example, they can be used by blind people to read information on any surface via vibration feedback similar to Braille [Ozi17]. Other use cases of tactile gloves are for example wayfinding [Zel03] or the transduction of visual stimuli to make games accessible for the blind [Yua08]. Handheld electrode grids were envisioned as tactile vision substitution system to perceive the environment [Kaj14]. Ultrasound feedback on the lips, teeth, and tongue can also convey virtual surface touch experiences [She22]. However, Shen et al. have so far focused on immersion and have not yet explored virtual texture sensations in the mouth region.

A promising approach to change the tactile perception of ubiquitous surfaces is *reverse electrovibration*. Instead of charging displays electrically with an alternating current [Bau10], a field is generated at the user's body that creates an *"oscillating electrical field around the user's fingers"* [Bau12]. When a touched surface is connected to the same ground as the user, tactile effects can be modulated on the surface. Bau and Poupyrev envisioned that their approach could be used in the future to help blind people orient themselves in public spaces. Furthermore, virtual textures can be rendered on physical objects to support AR applications. Another promising technology are active tattoos applied to the skin. When the tattoo is used to change the perception of a touched surface in the environment, it can be considered a passive surface interaction [Wit18]. In contrast, if tattoos are applied to the skin to detect input events, this can be considered an active surface interaction [Wei15].

Our gait is influenced by the perception of the ground [Pre73]. Each step produces a sound and haptic feedback which researchers have already proposed methods to manipulate e.g. to influence gait [Cor20; Gom20; Nor10]. Experiments by Tajadura-Jiménez et al. suggest that augmented footstep sounds also affect the self-perception of the body weight [Taj15]. Haptic feedback from shoes [Son18; Tak10] and also combined with augmented sounds [Nor10] were found to influence the perception of a surface. Those effects contribute to increase the perceived level of presence in virtual realities [Ker20] and also influence the user's gait in VR [Hop19]. In Chapter 9, we will take a more detailed look into the acoustic augmentation of footsteps. Again, we can find methods related to active and passive surface interactions. For example, if a floor is instrumented with step detection and auditory feedback from loudspeakers, we refer to it as an active surface [Ren21]. If the user wears headphones to perceive augmented footsteps in the wild, we refer to this as a passive surface interaction [Nor10]. While active surfaces require to instrument the environment, passive surfaces require mobile technology for interactions. Active surfaces have a limited interaction area and passive surface interactions enable to turn any environment into an interactive space.

2.5 Summary

In this chapter, we have shown that three classical senses are mainly involved in the perception of surfaces and textures. In addition to tactile feedback, visual and acoustic stimuli are responsible for giving us an overall impression that leads to the perception of a surface. We have shown how our senses transduce environmental stimuli into electrical signals and where in the brain the signals are mainly processed. However, we still do not fully understand the perceptual processes, although we have basic models describing the steps involved. This already enables us to influence perception of surfaces in many ways. Moreover, it is therefore reasonable that the origins of surface computing lie in the early history of human-computer interaction.

We continued by further defining surface interactions. Based on a condensed and brief presentation of related works, we could show that surface interactions can be divided into active surfaces and passive surfaces. It is noticeable that in many cases active surfaces can be converted into passive surface approaches and vice versa. An early example is the *DigitalDesk* by Wellner [Wel91] in 1991. Since cameras and projectors are installed in the environment, we can categorize his approach into active surfaces. Two years later, Feiner presented his augmented reality application for an office desk [Fei93]. While Wellner's vision is limited to a dedicated working space, Feiner's idea was more generalizeable to any office desk. However, at that time head-worn AR had not the same capabilities as projections.

Today we are much closer to the everyday use of augmented or even virtual spaces, although challenges still remain. For example, it is still the subject of research to make the tactile texture of virtual surfaces and the visual representation more realistic. It is also of further interest to explore the capabilities of passive surface interactions. Given their ubiquity, there are almost infinite possibilities to integrate ordinary passive surfaces into interactions with computational systems. The goal is to close the gap between the real and virtual worlds and enable a seamless transition between them. This means that not only the augmentation of real surfaces is taken into account, but also features of a physical surface are integrated into an interaction. Here, for example, we have mentioned digital pens that digitize handwriting while writing on a sheet of paper. Making physically bound information digitally available means that the information can be retrieved anywhere at the moment it is written.

Furthermore, challenges of passive surfaces can be identified. Active surfaces offer a more clear affordance to interact with technology. Passive surfaces, on the other hand, do not offer affordances for digital interactions. This means that the context of use must be recognized so that a computational system knows if any interaction is to be carried out. In addition, since every surface can be involved in an interaction, this could lead to an overload of superimposed stimuli. Various sensory systems are involved to interact with passive surfaces and to identify the context of use. Body-worn technology might raise privacy concerns for users and people around them. Social acceptance is necessary for seamless integration of passive surfaces into interactions [Bil15]. On the other hand, active surfaces are spatially bound in the physical space, require more resources in terms of rare earth materials for the required electronics in a ubiquitous context, and can also cause privacy concerns. In conclusion, it can be said that both subcategories of surface computing have their pros and cons and play an important role in human-computer interaction. The integration of passive surfaces in interactions with computational systems has opened up new applications with the advent of AR and VR headsets.

CHAPTER 3

Surface Scratching Sounds as Complementary Features in Handwriting

A specific case in which people interact with surfaces in everyday life is writing with a pen on paper. However, the written information is physically bound. This means that one must carry the paper with them to retrieve the information. Subsequent digitization makes the physical data digitally available everywhere but requires additional effort. This chapter is intended to investigate how scratching sounds of the pen tip on paper contribute to recognize handwritten digits. For this purpose our pen prototype *Pentelligence* is introduced, which measures the motion of the pen by an internal accelerometer and gyroscope as well was scratching sounds by an embedded contact microphone. We show an approach to integrate audios into the recognition of handwritten digits with motion of the pen without predefined writing instructions. Thus, the following research questions are addressed by this chapter.

> Do scratching sounds on paper provide complementary features to the movement of a pen that contribute to handwriting recognition?
>
> Does the combination of scratching sounds with the movement of a pen improve the recognition of handwritten digits?

This chapter is based on the master thesis by Max-Ludwig Stadler [Sta17] and a full paper at *CHI 2018* [Sch18] as well as a german patent [Max18a]:

- Schrapel, Maximilian & Stadler, Max-Ludwig & Rohs, Michael: ‘Pentelligence: Combining Pen Tip Motion and Writing Sounds for Handwritten Digit Recognition’. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2018: pp. 1–11
- Maximilian Schrapel, Luise aus der Fünten (EZN): *Pentelligence: Handschrifterkennung mittels Audio und Bewegungsdaten von Stiften*. DE Patent 10 2018 107 409 A1. 2018

3.1 Introduction

Traditional input methods like hardware keyboards are not suitable for small devices such as smart watches. Even on tablets software keyboards have the disadvantage that they overlay the display partially. An issue of direct touch input is the “fat finger” problem [Wig07], i.e., the occlusion of the target area under the finger. For precise input such as writing or drawing, capacitive touch pens and inductive stylus pens help to overcome some of these issues [Bra08]. In note taking tasks users often still prefer analog documents [Ste07]. This indicates a gap between the analog and digital world.

Digital or smart pens that operate on paper have the potential to create a bridge between both worlds. There are various products and technical approaches on the market. The *livescribe 3 pen* by *Anoto* [AB22] applies paper with a micro dot pattern and an infrared camera to detect strokes. Based on ultrasound and infrared light the *Inkling Pen* by *Wacom* [Com22] can operate together with a receiver box on any A4 paper sheet. Both systems can digitize handwriting and sketches. Digital pens are useful in many application scenarios. For instance, students can easily share and recall their notes from a lecture [Miu07] or patient charts can be directly digitized in hospital work [Zam07]. But the need for special paper (as with the Anoto pen) or constraints in writing space with receiver boxes (as with the Wacom Inkling pen) might be problematic in some situations.

In this chapter we present Pentelligence a novel digital pen, which uses a microphone and an inertial measuring unit to capture audio and motion data for handwriting recognition. Our prototype does not differ substantially from ordinary pens in appearance, weight, and size, and costs less than 10$. It combines the strengths of audio and motion data and achieves high recognition rates on our data set of about 9400 digit samples. The data set was taken from 26 individuals and classified by deep neural networks with majority voting.

Up to now, researchers focused on approaches with sensors such as cameras, motion or audio for handwriting recognition, but to the best of our knowledge we are the first to unite audio and motion on pens. We show that a combination of motion and audio can achieve better results than single sensor approaches. Furthermore this pen is not constrained to predefined digit trajectories, which is an important fact for the usability because of the individual writing style of every human. For this purpose the classifiers are retrained on a single writer. We also investigated in the acceptability of our prototype and point out future avenues of research.

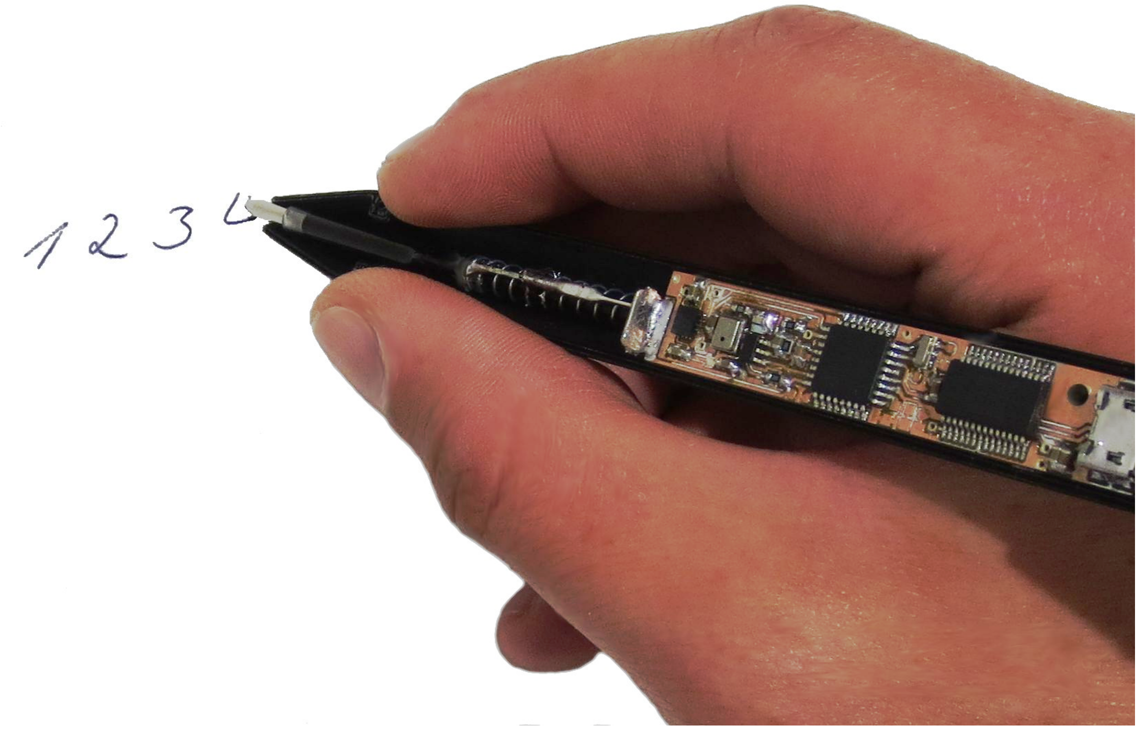

Figure 3.1: Handheld Pentelligence prototype while writing. The housing was removed for this picture to show the position of the inner hardware.

3.2 Related Work

Researchers focused on the field of handwriting recognition and gesture input for over five decades [Kri09; Liu03]. There are several camera-based approaches, which are sensitive to lighting conditions [Ara97; Bun99; Kis10; MIY02; Mun02; Nab95]. Our contribution is related to combining the strengths of motion and sound emissions from pens to achieve high recognition rates. Hence in the following this section is divided into captured motion and audio data for handwriting and pattern recognition.

3.2.1 Motion data for handwriting recognition

Using the motion of a pen for handwriting recognition is a well discovered field [Ban03; Bas08; Cho06; Wan12]. Applying inertial sensors for handwriting and gesture detection does not require an external reference [Ban03]. For this purpose, various methods and classifiers have been evaluated. Mostly characters are modeled as trajectories. For instance, Choi et al. [Cho06] used a triaxial accelerometer with principle components analysis (PCA) to reduce the feature vector size and computation time for a hidden Markov model (HMM) classifier. They also applied dynamic time warping algorithm (DTW) for the Arabic numerals '6' and '0' due to their similarity with the result of 100% and 90.8% recognition rates for writer-depended and in-depended data, respectively. Wang and Chuang [Wan12] proposed a trajectory recognition algorithm for a accelerometer based pen. Time- and frequency-domain features were extracted from an accelerometer and reduced by kernel-based class separability (KBCS) and linear discriminant analysis (LDA). A probabilistic neural network (PNN) on their data set achieved an overall recognition rate of 98% for handwritten digits and 98.75% on predefined gestures. Bashir and Kempf [Bas08] presented a novel pen device with an inertial measuring unit combined with pressure sensors for password entry. The data set contained in summary 10 different letters, numbers, and symbols. DTW was applied to each and the combination of all input channels. They could show that the combination of input channels achieves higher recognition rates of over 99% with an immensely reduced response time of less than 500 ms. Hence, sensing of motion and pressure promises good results for handwriting recognition. Moreover, Hwang et al. [Hwa12] presented a cheap pressure estimation touch pen which captured the audio generated by the pen tip when the stroked on a touch screen. For this purpose, the audio data was filtered and analyzed in frequency domain.

3.2.2 Acoustic data for handwriting recognition

Pen or finger strokes provide a very rich basis for acoustic features that can be applied to handwriting or gesture recognition. Harrison and Hudson [Har08b] used a stethoscope to detect distinct multi-part gestures composed of taps, lines, and circles and different surfaces. Li [Li04] proposed to use envelopes of audio signatures for handwriting authentication. He demonstrated that Hilbert envelopes of writing audios "caused by the frictions between rigid pen-nibs and paper surfaces" achieve a recognition rate of better than 75% with a straightforward multi-layer back propagation neural network. Li and Hammond [Li11] applied the mean amplitude and Mel-frequency cepstral coefficients (MFCC) features for template matching with DTW. They achieved an accuracy of 80% for 26 English characters constrained to a particular drawing order. Seniuk and Blostein [Sen09] focused on pen acoustic emissions by taping a microphone to the midpoint of the pen shaft. Two data sets consisting of 26 words and the cursive lower case alphabet from a single writer were recorded. Three classification algorithms have been evaluated: Signal subtraction, peak comparison, and scale space representation. Their preliminary results show that the classification with peak comparison reaches 95% on words and 70% on letters with signal subtraction.

To distinguish from related work, we apply sound and motion data from pens to combine the strengths of each sensor for digit classification. Our approach is not constrained to a particular drawing order. Hence majority voting neural networks generalize the individual writing styles.

3.3 Hardware Prototype

We decided to develop a lightweight digital pen with nearly the same look and feel as a regular ball pen. Figure 3.2 shows a detailed overview of the Pentelligence prototype. Due to the minimization of integrated circuits we achieved a length of 12 cm by only 11 g and 12 mm diameter. The 1 mm thick housing is printed with a Lulzbot 3D TAZ and ABS material. The ball point together with the ink cartridge and spring is taken out from a conventional low cost pen. The ATmega328p microcontroller [Fig. 3.2 mark 3] collects audio and motion data as well as detected force of the pen tip and sends them in frames to the computer via USB. The microcontroller runs at 16 MHz and contains an Arduino bootloader for easy programming.

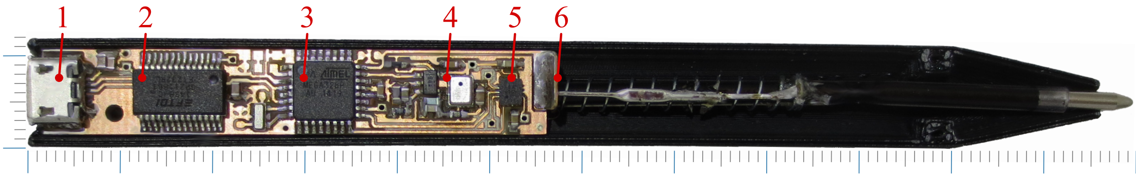

Figure 3.2: Detailed view of prototype in original size with markers for hardware explanations. The USB connector of the cable extends the prototype.
1 = micro-USB jack, 2 = USB/UART converter, 3 = microcontroller, 4 = microphone with amplifier, 5 = inertial measuring unit, 6 = write sensor.

3.3.1 Sensors

The segmentation of strokes and digits is done by detecting pressure on the pen tip. Hence, a contact sensor [Fig. 3.2 mark 6] is mounted on the spring to determine whenever a user writes on a surface. Moreover, by analyzing the write sensor first, computational overhead and unintended input can be avoided. The analog omnidirectional microphone MM34202 by DBUnlimited [Fig. 3.2 mark 4] is placed

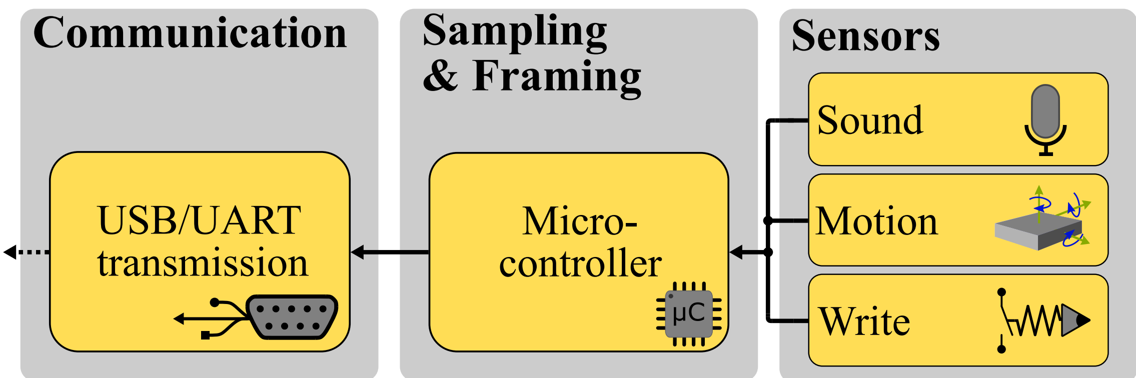

Figure 3.3: Simplified block diagram of hardware prototype.

on the top of the printed circuit board to avoid scratching audios from the ink cartridge on its housing. This MEMS microphone has a flat frequency response up to 16 kHz, a high signal to noise ratio (SNR) of 58 dB, and a sensitivity of -42 dB. The audio signal is amplified with an OPA344 rail-to-rail op-amp by Texas Instruments and measured with a 10-bit analog digital converter. The inertial measuring unit BMI160 by Bosch Sensortec [Fig. 3.2 mark 4], consisting of an accelerometer and a gyroscope with 16-bit values on each of the six axis, is placed as close to the pen tip as possible. Thus, it generates higher amplitude peaks when moved. All used parts are surface-mounted and mainly developed for smart phones and wearables.

3.3.2 Communication

The data of all sensors is sent to a computer, which is connected via USB [Fig. 3.2 mark 1], for further processing. Wired communication was chosen for the prototype because of its reliability and simplicity. In this iteration of the prototype we aimed to exclude potential problems with wireless transmissions. For this purpose UART was chosen because of its easy accessibility at the computer and availability in the microchip. The FT232RL USB-UART converter by FTDI [Fig. 3.2 mark 2] achieves in our implementation a data rate of 2 MBaud without errors. All sensor values are transmitted in a frame starting with an 8-bit sequence number for synchronization and frame loss detection at the receiver. In total each frame consists 128 data bits. The data rate is 7100 frames per second, which corresponds to 880.4 kbps. Figure 3.4 illustrates the structure of a frame and the frame extraction of the data stream at the PC.

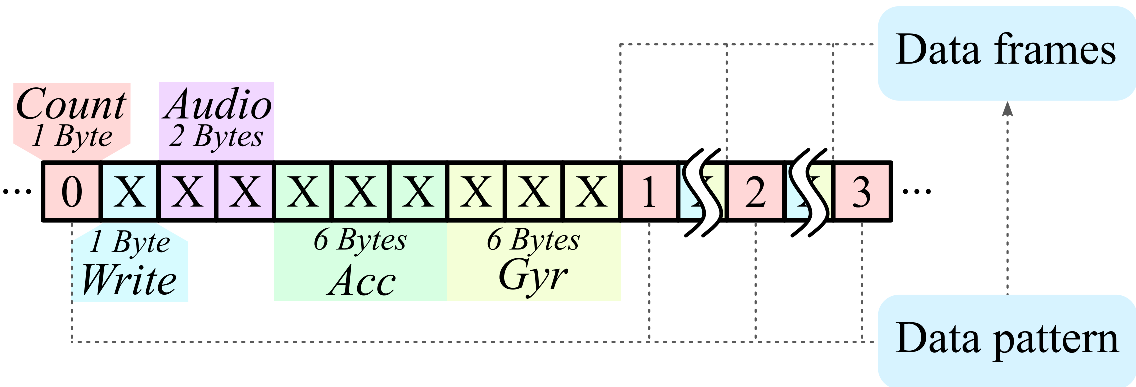

Figure 3.4: Illustration of the frame extraction based on a 8-bit sequence counter.

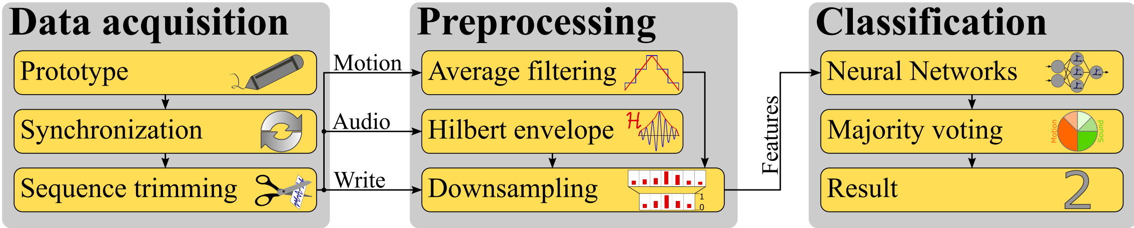

Figure 3.5: Block diagram of signal processing, partitioned in data acquisition, preprocessing, and classification.

3.4 Signal Preprocessing

A universal asynchronous receiver-transmitter (UART) splits all data to be sent into bytes with a start and stop bit [Osb80]. Hence all frames have to be re-assembled at the computer for which we use a counting 8-bit sequence number indicating the start of each frame as visualized in Figure 3.4. When the amount of bytes of a few frames are received the counting pattern can be detected and utilized for separation. The included data is then unpacked and stored in a buffer for further processing. The continuous change of the sensor values allows data extraction based on the sequence counter after at least three received frames. The reliability of the extraction can be further improved by using more frames for analysis. In addition, the sequence counter enables to detect packet loss and jitter. Since a wired connection on a cable length of 2 m is used, none of these problems occur. However, for future wireless iterations of the prototype, this method can be reused.

3.4.1 Write Sensor

The space between two written symbols on a piece of paper requires that the user move his hand to the new position. This space constraint causes a short period of time where the pen tip has no contact to the surface. Hence every symbol in the buffer is separated by a time period where no force can be detected. A preliminary test with two participants showed that 700 ms is an appropriate value for trimming the sequence. To exemplify the impact of sequence trimming the smoothed audio of a handwritten '7' is analyzed in Figure 3.6. The red lines beside of marker 5 represent the short time period of approximately 100 ms when the pen tip is lifted to write the middle stroke. Moreover by pointing the pen tip on a surface a high audio peak at markers 1 and 6 can be observed. This information could also be utilized for trimming the sequence but it is less reliable because grabbing a pen or putting it on the table would also generate similar acoustic emissions.

3.4.2 Audio processing

By further analyzing the smoothed audio sequence in Figure 3.6 on the left side it is to be stated that every stroke has its own acoustic features. At marker 3 the pen tip is not lifted from the surface but shows high signal peaks when the stroke direction is changed. During every stroke the acoustic emissions seem to be modulated on a resonance frequency of the prototype and surface. Hence it is necessary to analyze the raw audio stream in frequency domain to determine the range that contains the most information on handwritings. For this reason, each digit was written and directly transmitted to the audio card.

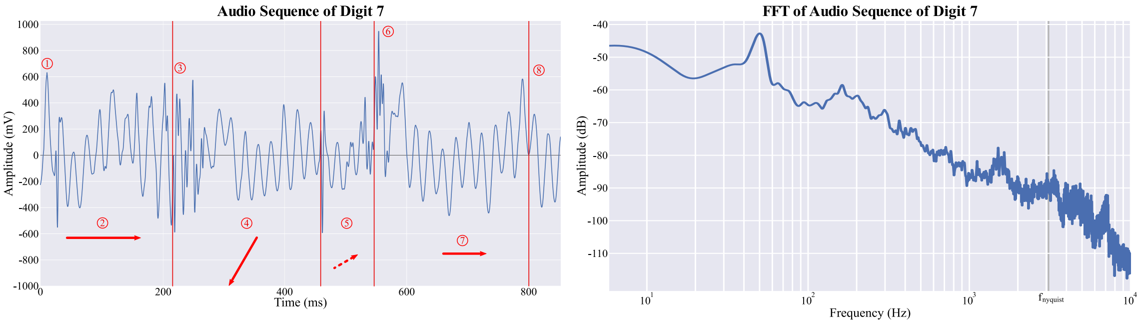

Figure 3.6: Smoothed audio sequence of a handwritten digit ‘7’ with explanations and related FFT with Hanning window below recorded at 44.1 kHz sampling rate with audacity. 1 = sound emission peak generated when pen tip touches surface, 2 = trajectory of the stroke, 3 = end of the trajectory, 4 = stroke downwards, 5 = re-positioning of the lifted pen tip, 6 = sound emission peak as 1, 7 = trajectory of middle stroke, 8 = pen tip is lifted and digit is finished.

The signals and its components were then analyzed with the audio editing tool audacity and a Fast Fourier Transformation (FFT) was applied. This approximation of the signal in frequency domain has the disadvantage of producing additional components by periodically and infinitely repeating the signal in a window when the signal is not starting and ending with 0 values. This so called leakage effect can be reduced by applying special windows such as Hanning which have different weights on the borders [Har78]. Figure 3.6 on the right demonstrates that most of the information of the handwritten digit ‘7’ lies below 2 kHz. This frequency distribution is produced by the structure of normal paper on a wooden table and the speed of stroking the pen tip. Consequently, our sample rate of 7.1 kHz, is sufficient to cover the frequencies that are relevant for classification.

Noise is another relevant aspect for pattern recognition with audio. Li [Li04] proposed to use the normalized Hilbert envelope of writing audios as the feature space. We compared this approach to an FFT representation (with a Hamming window) and to the raw audio data of each symbol. In an initial evaluation we recorded 40 samples of each digit from a single writer. All recorded sound emissions of each digit are then compared to all others with FastDTW [Sal07]. This algorithm provides a linear and accurate approximation of dynamic time warping. The procedure was repeated for Hilbert and Fast Fourier Transformation. The out-coming distances were averaged and compared to each other by normalizing the highest distance to 100%.

Table 3.1: Comparison of written digit dissimilarities from sound emissions. The resulting distances with FastDTW algorithm of every symbol are averaged and normalized to the method with the highest distance.

Dissimilarity comparison (normalized)	
Method	*Result (%)*
Hilbert	100
FFT	72
Raw audio	71.3

Table 3.1 shows that the normalized Hilbert envelopes clearly outperforms FFT and raw audio data approaches for a single writer. Recorded sound emissions by pen tips contain noise related to the surface as well as other sources such as resonances of the prototype itself. Li [Li04] stated that the movement of a pen tip on the paper imposes an modulation effect on that carriers. Hilbert envelopes help to overcome such issues by smoothing the signal course. Moreover determining and reducing the signal to the significant audio features of digits in preprocessing indicates better results for classification. After applying Hilbert envelopes on the incoming signal it has to be stretched to the number of input neurons because of the varying input length and normalized to a interval between 0 and 1 for internal processing of the classifier.

3.4.3 Motion data processing

The accelerometer and gyroscope values of each axis are down-sampled and averaged to 150 Hz. Due to the relatively slow motion of the pen tip, stretching the data to 150 input neurons is adequate for classification. For comparison, Krishnan et al. [Kri09] used a sampling rate of 100 Hz and achieved recognition accuracies of over 86% for 5 gestures with an Adaboost classifier. Wang et al. [Wan12] achieved accuracies of 97% for digits by using a sample rate of 100 Hz for their accelerometer-based pen with a probabilistic neural network classifier.

3.5 Classification

Handwriting itself is very complex and shows strong individual characteristics [Zha03]. We also observed that single participants of our study wrote same digits differently. For instance the number ‘5’ visualized in Figure 3.7 was often written in two styles by the same participants. Hence classification with template matching approaches may suffer under different trajectories and cannot detect both in a single step. Moreover the growing number of symbol variations would increase the classification time rapidly. Another important fact is the challenging feature selection for audio since this sensor provides very noisy data. Motion can be also challenging due to the rotation variance and the individual style of holding the pen. Thus, we decided to use neural networks for classifying handwritten digits. Furthermore this provides the opportunity of individualization for every user by retraining the network after collecting a few samples of each digit.

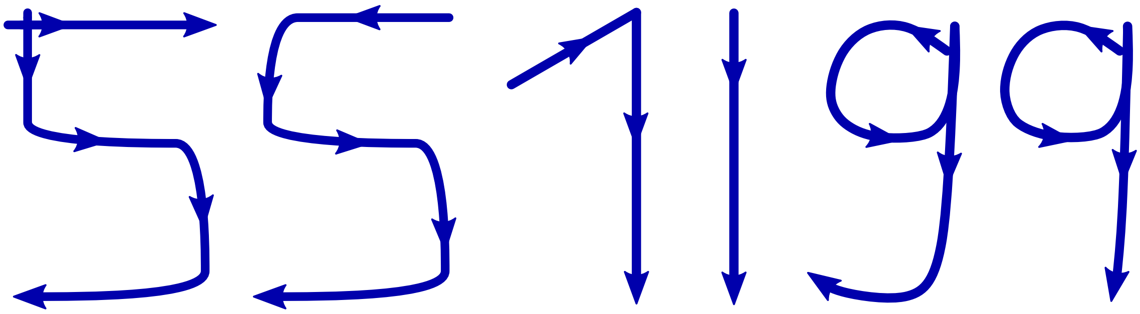

Figure 3.7: Left: different observed writing styles of '5' written by a single person. Middle and right: typical trajectories of '1' and '9' written by different persons.

3.5.1 Dataset

We collected a dataset to capture digit samples in various writing styles. 23 male and 3 female volunteers (average 26.54 years, SD 8.038) were recruited including three left handed individuals. To the sound of a metronome with 30 beats per minute every participant was instructed to write one digit repeatedly for 65 seconds on a DIN A4 squared paper when the click was heard. The study was conducted in a quiet environment and the metronome sound was set that quiet that the participant could still hear it. Hence problems with trimming the sequence could be avoided as well as noise influences due to the environment. The housing and the relative loud sound emissions of the scratching sounds prevent from impacts of metronome audios on the dataset. A random order of repeatedly writing digits was chosen to eliminate problems with frustration because this influences the pressure and speed of a user moving the pen tip [Asa13]. Furthermore this leads to different sound emissions on a surface [Hwa12]. In addition 5 male participants did the study twice to collect data for the test set and one of them made additionally 212 test samples of each digit for individualization of the classifier.

In summary 6169 training and 3239 test samples were collected. The individualization set contains 2120 samples divided into 90 training and 122 test samples of each digit. During the study people may missed a click of the metronome. Hence the amount of collected samples decreased. After completing the study every participant was handed out a short questionnaire to give feedback on the usability of the pen.

3.5.2 Neural Network Configurations

Finding an optimal configuration for deep neural networks is a challenging task, because many parameters have to be considered. We built classifiers for motion data with binary writing information, and for audio alone as well as in combination with motion and writing data. To proof the statement that binary writing features together with audio does not improve the accuracy in comparison to audio alone we also implemented a neural network with both sensors.

Determining the optimal hyperparameters, e.g. number of hidden layers and neurons for all classifiers is a very important part. If the complexity of the networks are too low they achieve high error rates on the dataset. If the complexity is too high long training and recalling times are the consequence. Hence many rules of thumb and optimization algorithms have been purposed over the last two decades [Vuj16]. Our specific problem is optimized by trial and error on different network topologies with the aim to apply not more than 5000 training iterations. The results in Table 3.2 indicate that audio provides more difficult features than motion because 4 hidden layers have to be applied which corresponds to a very abstract signal representation at the fourth hidden layer.

Table 3.2: Applied number of layers and neurons of neural networks using motion with write information, audio alone and together with write information and all sensors together.

Neural Network topologies

Layer		**Network** *Motion & Write*	*Audio*	*Audio & Write*	*All*	
Input		1050	3500	3650	4400	**Neurons**
hidden	*1*	1050	3500	3650	4400	
	2	790	2802	2952	3442	
	3	530	2104	2254	2484	
	4		1406	1556	1406	
Output		10	10	10	10	

Beside of the optimal network topologies in Table 3.2 overfitting of the classifier on the training set has to be considered. When the networks co-adapting too much during training procedure, they loose accuracy on unknown samples. Dropout, a technique to avoid overfitting by temporarily dropping a set of randomly selected neurons and their connections, is applied [Sri14].

Achieving a high generalization in one neural network is challenging due to the different observed writing styles in our dataset. The complexity of the optimal topology is maybe too high to be trained on our training samples. Alpaydin proposed to train a number of neural networks independently and apply a voting scheme which increases the generalization significantly [Alp92]. Thus majority voting networks with the same topology and voting weight are examined. For this purpose the training set is divided by the amount of participating networks into sub-sets and the accuracy is compared to single classifiers. To bind the strengths of audio and motion a combination of both voting classifiers is also observed.

3.5.3 Individualization

To achieve higher accuracies the classifiers are optimized to the individual writing style of a target user by applying a second training step. The individualization dataset from a single writer is utilized to evaluate this approach. The neural networks for audio as well as motion with writing information are retrained and the overall and single digit accuracies are examined alone and in combination of both majority voting classifiers.

3.6 Results

The classifiers are evaluated by the following results beginning from examining dropout rates on our dataset following with single and majority voting neural networks. Further the individualization for a single writer is examined as well as the results of a short questionnaire about the prototype.

3.6.1 Dropout

The first point of interest is the dropout rate for the neural networks of our approach. Table 3.2 shows that the highest accuracy can be achieved when a dropout rate of 25% is applied for acoustic emissions from handwritten digits. This result was also confirmed by the other neuronal networks from Table 3.3. Reducing or increasing the percentage of randomly omitted neurons from each layer during training iterations leads to lower accuracies. Hence the further examinations are related to a dropout rate of 25% in training procedure of all classifiers.

Table 3.3: Comparison of dropout rates for single neural networks trained with audio data. In every training epoch a percentage of randomly chosen neurons from each layer is omitted to avoid overfitting.

Dropout	*Accuracy*
0%	0.70
10%	0.74
25%	0.79
50%	0.76

3.6.2 Classifiers

By applying a dropout rate of 25% the influence of the number of networks on the accuracy is to be examined. For this purpose, a single neuronal network is compared to 4, 5, and 8 majority voting classifiers with the same setup.

Table 3.4: Precision and Recall of classifiers with various amount of voting neural networks in same topology.

		Precision	*Recall*
Nets	1	0.58	0.57
	4	0.61	0.6
	5	0.60	0.59
	8	0.52	0.52

The results in Table 3.4 show that majority voting with four audio classifiers outperform single neural networks. All other topologies from Table 3.2 validated these findings. Moreover a one-way-ANOVA was conducted and confirmed statistically that single networks achieve lower accuracies than four majority classifiers on or dataset (F_{dfs}=7.11, $p<0.01$). Hence, the further examination is based on majority voting among four networks.

To observe the individual performance of each test user on the different classifiers Figure 3.8 shows the mean precisions on all digits. The test users 1 to 5 conducted the study twice to generate samples for the train and test dataset. Users 6 to 10 are completely unknown to the classifiers and achieve lower accuracies on all neural networks. User 9 was the only left-handed person in our test dataset and shows in comparison to all others the lowest recognition rates. Moreover his individual result is the only where audio performs better than motion. On all other users motion data achieves the highest precision. The write sensor information together with audio performs slightly better on known writing styles than audio alone. When the user is not known to the classifier audio alone shows better results on right-handed persons of our training dataset with less left-handed individuals. Networks with all sensors achieve a slightly higher precision than audio alone when the writing styles are known for the classifiers.

To point out the strengths of acoustic emissions and motion from pen tips, the accuracy of single digits on our dataset has to be observed. Figure 3.9 compares confusion matrices of audio alone to motion with write information and all sensor data in one classifier with each other. In comparison of the classifiers audio achieves an overall accuracy of 58.1% while all sensors together reach 60.6% and motion with writing information outperforms both at 79.2%. Handwritten digits ‘0’ are often misclassified by networks with motion and write information as ‘6’ while audio classifiers confuse them less frequently when the digit is actually a ‘0’. A particular observation is that motion with write data classifiers have a lower accuracy on digits ‘5’ and ‘9’ than on others. A ‘9’ is often classified as a ‘3’ while vice versa 95% of all samples are predicted correctly. For this specific case audio networks have a higher accuracy. A similar phenomenon can be found for the digits ‘5’ and ‘7’.

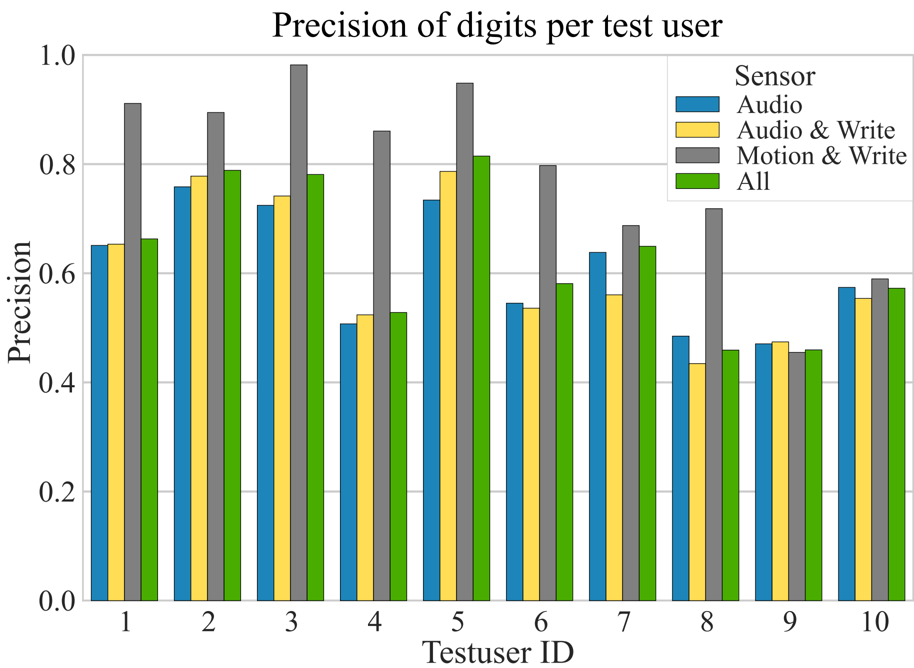

Figure 3.8: Mean Precision of digits for all test users with various classifiers. The majority voting neural networks were trained with different sensor combinations. Users 1 to 5 are known while 6 to 10 are completely new for the classifiers.

The probability of predicting a '5' when it is actually a '7' is 16.3% while the audio classifier has a probability of only 10.8%. From the findings in Figure 3.9 and 3.8 it can be stated that using all sensors in one majority voting classifier is not as applicable as audio or motion with writing information alone. The networks could not identify the strengths of each sensor to achieve higher accuracies than motion classifiers. Hence, the found strengths have to be generalized and applied differently.

For this purpose, motion first classifies an unknown digit and when the outcome has a higher accuracy on audio networks they verify the result. For instance if a '3' was predicted by the motion classifier, audio networks proof the decision by determining that it is not actually a '9'. This exchange of accuracies in the confusion matrices causes a theoretically increasing accuracy on handwritten digits '9' to 76.2% with a decreasing accuracy on digits '3' to 78.3%. In comparison to the confusion matrix of the motion with write information classifier in Figure 3.9 the overall accuracy decreased theoretically to about 78.86% with the advantage of a better precision on handwritten digits '9' while decreasing the accuracy on digit '3'. Re-evaluation of a predicted '6' that it is not actually a '0' can also increase the overall accuracy to 79.5% by a decreasing precision on the digit '6'.

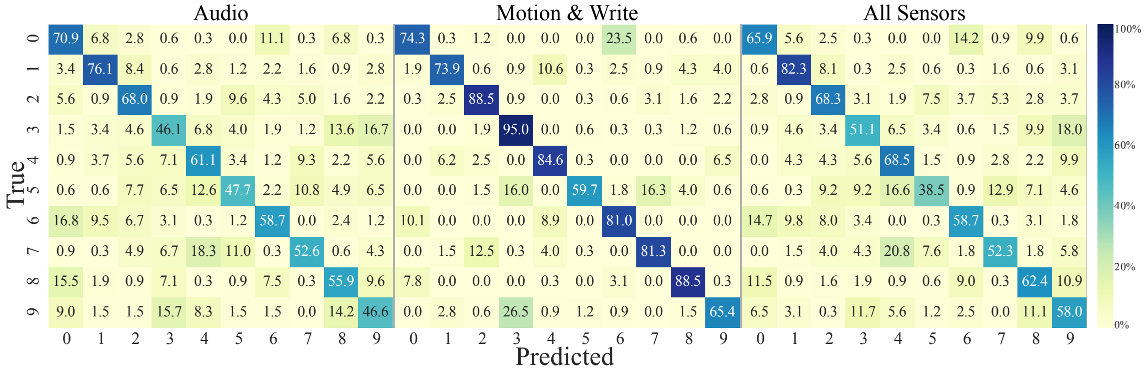

Figure 3.9: Relative confusion matrices of classifiers trained with audio, motion & write information and all sensors together.

To generalize this exchanging rule let Mat_M be the confusion matrix for motion with write information and Mat_A for audio data. Further the indicator t is the true and f the false predicted digit for exchange. If the rows and columns of the confusion matrices in Figure 3.9 are normalized to the interval [0,1] then the distance $d_{g_{t,f}}$ in equation 3.1 is the increasing accuracy of the target digit and $d_{l_{t,f}}$ in equation 3.2 the decreasing accuracy of the false predicted digit. If the exchange with the audio classifier increase the mean accuracy of the true and false predicted digits in equation 3.3 then the re-validation rule can be applied. To avoid the classifier from re-evaluating too much, a threshold $\varepsilon \geq 0$ can be applied and the rule can only be utilized once per digit for not decreasing the accuracy on single exchange digits too much. Furthermore by permitting negative thresholds for the worst classified digits by motion data the confusion can be attenuated.

$$d_{g_{t,f}} = Mat_M[t,f] - (Mat_M[t,f] \cdot (1 - Mat_A[t,t])) \tag{3.1}$$

$$d_{l_{t,f}} = Mat_M[f,f] - (Mat_M[f,f] \cdot (1 - Mat_A[f,t])) \tag{3.2}$$

$$d_{g_{t,f}} - d_{l_{t,f}} \geq \varepsilon \tag{3.3}$$

If the found rule with a threshold $\varepsilon = 0.03$ is applied on the confusion matrices in Figure 3.9, the classifier re-evaluates the result of the motion networks between the digits '0' to '6', '5' to '3' and '1' to '4' or '6' to '4' on audio data. To increase the accuracies of the most confused digits '5' and '9' negative thresholds are permitted for these symbols. Hence the exchange rule is applied on '0' to '6', '3' to '9' and '5' to '7'.

The theoretical combination of both classifiers can not state the real results because the statistical distribution of the test samples is unknown. Hence the resulting combined classifier of all sensors was tested on our dataset. The theoretical rule for combining the strengths of motion and audio could be confirmed by the confusion matrix in Figure 3.10. The features of captured audios and motion of pen tips provide complementary features which can be combined for increasing the accuracy on single handwritten digits. The individual writing styles in our dataset can not be generalized to achieve applicable accuracies for human-computer interaction. Hence, the individualization of the classifiers is to be observed.

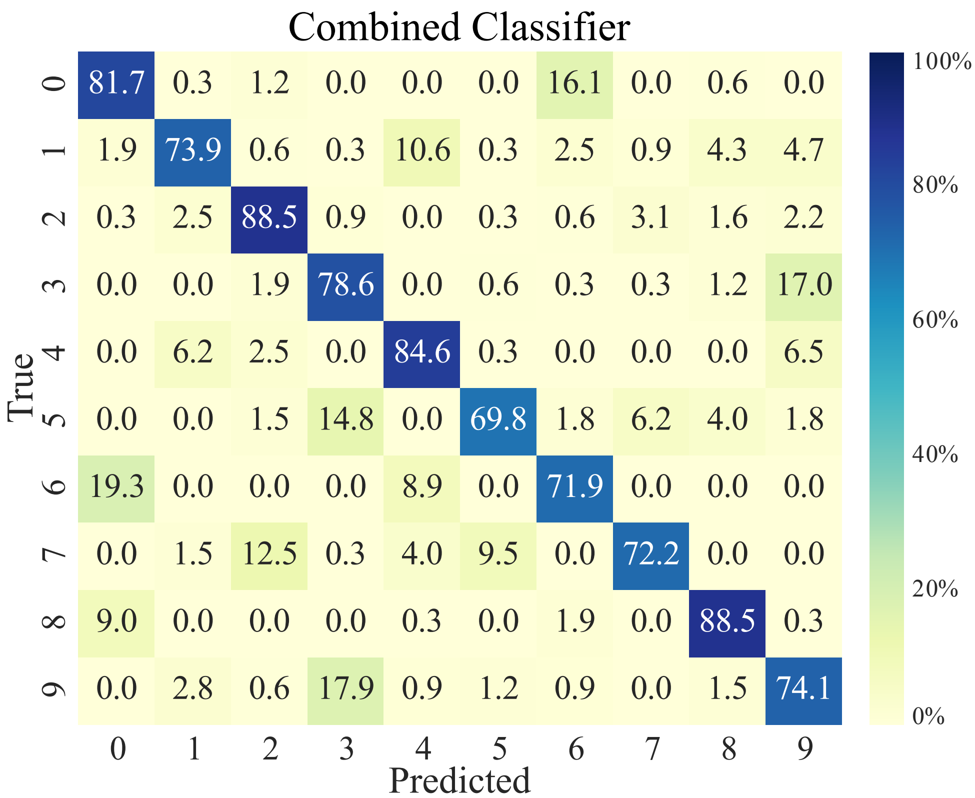

Figure 3.10: Confusion matrix of combined classifier. First the samples were predicted with motion and write sensor data, then the outcome was validated by the audio classifier.

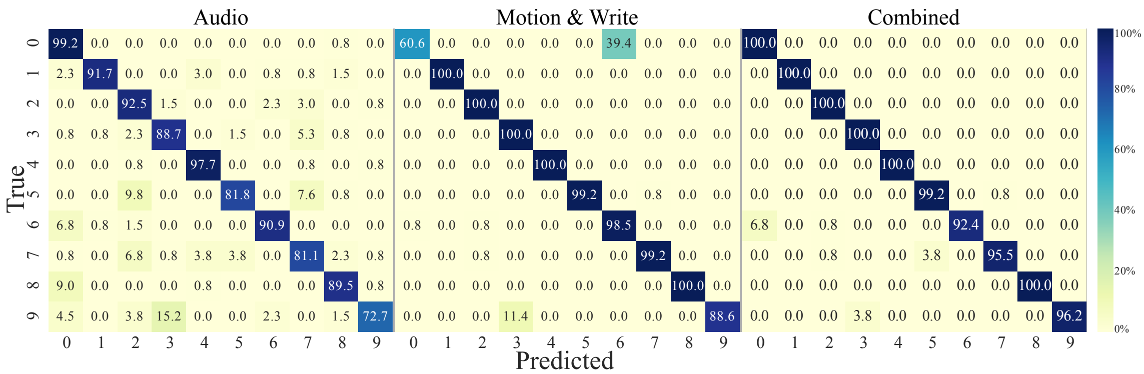

Figure 3.11: Relative confusion matrices of audio, motion & write information and combined classifiers with re-validation rule for a single writer.

3.6.3 Individualization

The classifiers were retrained with 90 samples of each digit by a single writer. For this purpose, 50 training iterations were chosen. The previously defined exchanges between the digits '0' to '6', '3' to '9' and '5' to '7' were applied for the individualized combined classifier.

The results in Figure 3.11 for a single right-handed writer confirm that audio and motion of pen tips have complementary features which can be utilized to achieve high recognition rates. Especially handwritten digits '0' are often confused by motion data with '6' while acoustic emissions do not show any errors on that case. The audio classifier achieved an overall accuracy of 88.58% while motion with writing information reaches 94.61%. In combination of both very high recognition rates of 98.33% outperform single classifiers. Moreover, all test samples of '0' to '4' and '8' are 100% correctly predicted which corresponds to 722 test samples. The worst result of 92.4% on handwritten digits '6' is related to the re-validation by the audio classifier and confusion to digit '0'. An exchange of the digits '3' to '9' achieves even higher accuracies in that individual case because of complementary features of motion and acoustic emissions from pen tips. The overall and single digit accuracies of the combined classifier state that the individual writing style of a single writer can be generalized.

3.6.4 Questionnaire

After completing the study, every participant was handed out a short questionnaire. The results are shown in Figure 3.12.

At first we wanted to find out if people are concerned about taking samples from their handwriting. This is necessary to know because the initial dataset had to be collected. Only 16.6% of all surveyed participants were worried about privacy issues. The same amount of participants agreed that they don't see any privacy issues. Over

54.1% strongly agreed that they don't mind about storing audio and motion data from their handwriting.

We also wanted to know if our participants do note-taking in their everyday life because this indicates the applicability of our prototype in comparison to other digital pens with special writing surfaces on which the pen tip is tracked. In case of note-taking special point paper is maybe not available or receiver boxes have to be clipped on the paper, which is maybe in case of sticky notes too small. About 8.3% do not take notes while over 20.8% stated that note-taking is no part of their daily life. On the other hand 16.6% strongly agree and 37.5% declared that they do take notes but not in a very frequently daily fashion.

For further improvement of our prototype we also had the question if the wired connection bothered the participants. Over 45.8% decided that the cable has to be removed for not disturbing them while writing. About 16.6% were not bothered by the cable and the same amount of participants neither agreed nor disagreed.

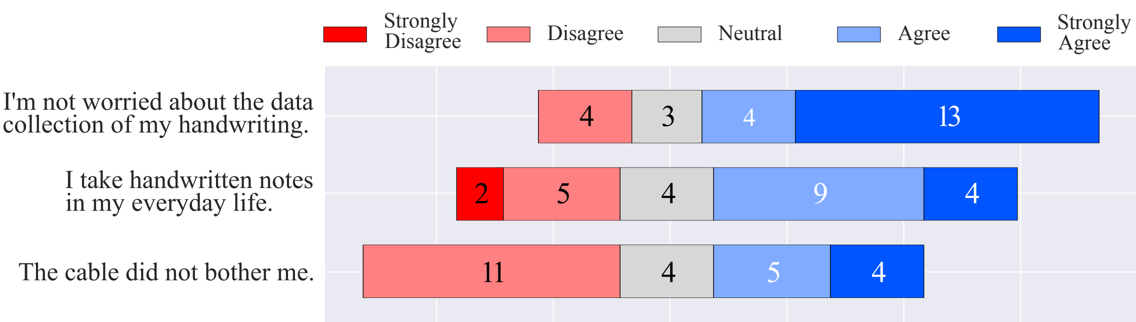

Figure 3.12: Results of the questionnaire on 24 participants in the study handed out after completing the data acquisition.

3.7 Discussion

Regarding the hardware prototype and questionnaire, the cable should to be replaced by a wireless connection. Our claim is to avoid any constraints in writing space and for that purpose cables are not applicable. Reviewing note-taking as a field of application, so far this pen can digitize information such as phone numbers or be used as calculator. In addition handwritten letters have to be considered. Furthermore 16.6% were worried about collecting samples of their handwriting, which is may related to assembling their signature and faking identity.

By reviewing the results of the classifiers it can be stated that our dataset provides a rich base of different handwriting samples. Overfitting is a observed problem which can be reduced by applying dropout during training process. Moreover the individual writing styles can not be generalized by one classifier. Applying majority voting with four neural networks achieves the highest precision on every sensor combination.

Acoustic emissions produced by pen tips scratching on paper have more confusion between digits than corresponding motion measurements. This is may founded in surface related features and audio back coupling paths of the prototype. Moreover audio streams provide less features than 6-axis motion sensors. Hence the signal has to be classified in more hidden layers to achieve acceptable results.

Motion features are more reliable than audio. It can be stated that motion classifiers have less confusion between all digits but were less accurate when, for example, the input is predicted as ‘6’ but it was actually a ‘0’. This is founded in the similar trajectories and writing styles of both digits. It is challenging for the motion classifier to differentiate both especially when no writing instructions are given. In such specific cases audio provides complementary features which can be utilized to improve the accuracy.

When all sensors are applied in a single classifier with majority voting neural networks generalization is also challenging. The significant features of each sensor are not identified by the classifier to achieve higher recognition rates than motion alone. The confusion over all digits of acoustic emissions strongly influence the classifier and decrease the accuracy. Only in case of handwritten digits ‘1’ where audio features have a higher accuracy than motion the classifier can combine both strengths. The input weights of audio features could be decreased to achieve a higher focus on motion data. Furthermore the number of input neurons for audio data could be decreased but an example FFT analysis of an audio stream has shown that this would result in a loss of signal features.

To overcome challenges in combining the strengths of each sensor a re-validation rule was presented which evaluates the result of motion data by classifying the audio stream. The results show that the accuracy on digits can be increased by utilizing the complementary audio features. Moreover for this purpose the accuracy on the confused digits had to be decreased but there are also cases on digit such as ‘0’ and

‘6’ where the overall accuracy compared to motion classifiers can be increased.

The achieved accuracy of 78.38% is not accurate enough to fully utilize the prototype for digitization tasks. It can also be stated that the mean precision on digits varies between all test users. Sound emissions provide rotation invariant features which can be an advantage for left-handed user if the network was previously trained on more right-handed users. In that specific examined test case the accuracy of less than 50% on rotation sensitive motion data is below audio. This is founded in the observed different way of holding the pen by left-handed persons. Even right-handed people show various holding styles. One right-handed test user achieved accuracies below 60%. All individual styles can not be generalized by our dataset.

Hence, an individualization of the classifiers was applied. A single writer wrote 90 samples of each digit and the networks were re-trained for 50 epochs on the individual dataset. The challenges in confusing the digits ‘0’ and ‘6’ by motion classifiers increased while acoustic emissions overcome such problems. Combining the complementary features by the previous defined rules increased the overall accuracy on handwritten digits for a single writer to 98.33%. Hence this also confirmed that the individuality of handwritten digits can be generalized for a single writer even when there is a variance in writing single digits.

By providing previously trained networks on different writers, the amount of handwritten digits and computing time for training the classifiers on a target user can be reduced. This procedure has only to be done once and could also be trained on other surfaces such as hand held clipboards to provide a higher reliability for specific use cases.

3.8 Conclusion and Future Work

We reported the design of a digital pen for common writing surfaces, such as paper, that does not require a special pattern and is not confined to a particular area. The designed pen does not strongly differ from regular pens in appearance (except for a wired USB connection) and is build from cheap and robust components.

The evaluation based on a corpus of digit data shows that the pen achieves high recognition rates. The combination of motion and audio data performs substantially better than motion data alone or audio data alone. It turns out that these types of sensor data have complementary characteristics. Thus, we can conclude that acoustic emissions from friction of the pen tip on the paper contribute to recognizing handwritten digits.

It was also shown that majority voting neural networks with the same topology provide a more accurate classification than single networks. Furthermore it came out that combining the strengths of audio and motion features achieves better results when first motion data are classified and then the outcome is evaluated by the acoustic emissions of pen tips. Retraining the classifiers for a single writer on a dataset of 900 samples turned out to achieve very high recognition rates of 98.33%. It was also stated from a survey that users accept the recording of their handwriting for individualizing the neural networks.

For testing the prototype in real world applications the wired connection should be replaced by a wireless version according to the results of the questionnaire. To provide a wide field of applications the dataset has to be extended to letters and gestures. Deeper neural networks with a higher amount of training data and different topologies could be observed as well as the impact of different surfaces or background sound. A further examination of the impact of right and left-handed users on the accuracy of each sensor is also a point of interest. A live application such as phone number dialing can be implemented to evaluate the user experience. Writer identification is also a research topic where this pen could be applied.

One limitation of the approach is that it is not possible to determine the writing position on a sheet of paper. Hence, sketching is not supported. Moreover, samples of handwriting must first be recorded before high accuracies can be reached. This labeling process can be perceived as exhausting and frustrating. However, 90 samples of each digits were already sufficient to achieve high recognition rates. High accuracies are necessary because, unlike written words, no subsequent auto correction can be applied. For letters, incorrectly recognized symbols that are subsequently corrected could be automatically added to the training set for individualization.

CHAPTER 4

Individuality of Surface Scratching Sounds in Biometric Writer Identification

In the previous chapter, we found that scratching sounds produced by the pen tip when moving on paper contribute to recognize handwritings because they provide complementary features to the motion of a pen. This opens up the question of whether these scratching sounds also include user-specific features that, in combination with motion data, are also capable of identifying the writer. For this purpose, the previously introduced prototype *Pentelligence* is used with a more ergonomic housing. We analyze whether the scratching sounds contribute to identify the writer by the use case of symbolic signatures. The presented application scenario thus implements an online signature verification in a particularly challenging identification task. We specify writing instructions for the symbols in order to make the movement of the pen when stroking the pen tip on the paper similar among writers. This serves to answer the following research questions:

> Do scratching sounds on the paper contain information for identifying writers?
>
> Can audio and motion data from digital pens be combined to identify the writer?

This chapter is based on the bachelor thesis by Dennis Grannemann [Gra18] and a full paper at *MuC 2022* [Sch22a]. Dennis collected the dataset during his thesis which was subsequently analyzed by Maximilian Schrapel on the applicability to identify writers. In addition, the presented preprocessing steps were first tested in the bachelor's thesis of Lukas Nagel [Nag18].

- Schrapel, Maximilian et al.: 'Sign H3re: Symbol and X-Mark Writer Identification Using Audio and Motion Data from a Digital Pen'. *Proceedings of Mensch Und Computer 2022*. MuC '22. Darmstadt, Germany: Association for Computing Machinery, 2022: pp. 209–218

4.1 Introduction

Even today, handwritten signatures are an important means of identifying a person. Although electronic signatures offer a digital alternative [Doc21], the legal situation defines several cases in which a physical handwritten signature with a pen on paper is necessary. According to the National Telecommunications and Information Administration (NITA), cases include, for example, wills and testamentary dispositions, family law cases (e.g. divorce or adoption), insurance policy cancellations, products with significant health risks, all documents for the transport of hazardous materials and utility cancellations [Tel02]. German law requires that every document in which rights are exercised must be signed. However, unless otherwise specified (e.g. testamentary dispositions [ges22b]), documents can be signed digitally [ges22a]. In addition, the finance and business sectors remain subject to physical signature requirements. Even today, businessmen travel around the world for notarized signatures in order to conclude valid contracts. In the financial sector, paper checks and debit cards require signatures for validity. Decentralized finance (DeFi) aims to offer a future alternative [Eth21a]. For instance, in contrast to banks, smart contracts on the Ethereum blockchain are not subject to human involvement. They can be seen as a type of account in the network with a balance and functions defined by code. Other accounts can irreversibly submit transactions to execute these functions [Eth21b]. Smart contracts offer various advantages, including anonymity, efficiency, transparency, and providing financial services to people without access to bank accounts [Eth21a]. However, legal enforceability is still unclear [Gia17], implying that signatures are likely to remain important. Handwriting is unique because it is influenced by training, physiology, and further behavioral factors over a lifetime [Gam80; Sri08]. Verifying the validity of a signature poses a major challenge. The mismatch of signatures on mail ballots with signatures on file led to the rejection of thousands of votes in the past 2012 and 2016 U.S. presidential elections [Smi18]. In the same time period cases of signature fraud in Canada tripled between 2014 and 2016 [Eri17].

These examples demonstrate the need for security precautions due to legal requirements and the risk of forged signatures. The verification of x-mark signatures in particular, which must be witnessed for legal validity, presents a major challenge. Illiteracy or disabilities may result in an individual not being able to write his or her full name. The witness of a x-mark signature serves to verify and confirm the identity of the writer. Hence, x-mark signatures do not provide sufficient security and validity for online signed contracts when no witness is involved. In cases of suspected fraud, witnesses of an x-mark signature must be investigated. Forensic methods on the x-mark handwriting are limited to stroke features like length, direction, orientation and connection as well as features related to pen and paper, such as line quality and the crossing point position [Hil92; Hub99; Mar18].

In this chapter, we present an approach to measure the individuality of handwritten symbols with predefined writing instructions for user identification. Recorded data of a digital pen when signing a contract can support in cases of suspected fraud to identify the writer. We use measured motion data and scratching sounds from a digital pen [Sch18]. We conducted a user study with 30 participants as a proof-of-concept and to collected a dataset of 19 different symbols for authentication. On this dataset different feature sets, which are fed into SVMs, are analyzed for their applicability for writer identification. In our tests, we focus on small training sample sizes, as it would be inconvenient for a person to write his or her x-mark signature several times before the proposed system is able to identify the handwriting. We show that by combining audio data – of the scratching sound that the pen tip makes when it moves across the paper – with 3D acceleration and gyroscope data of the pen, just three training samples of a symbol are sufficient to reach an average F1-score of over 87 % on writing three test symbols. In addition, alternative symbols were identified that are easy to draw and, with our approach, reach higher F1-scores than commonly used x-marks.

4.2 Related Work

Signature verification can be divided into offline [Sta21] and online approaches [Bib20]. Offline verification usually requires signature image analysis for proofing one's identity [Haf17; Lev18]. Therefore, it only relies on features from the signature image. Skilled forgers imitate the style of a signature to claim a person's identity. Online approaches use data recordings of the writing process to identify user-specific features. The larger number of features compared to offline approaches usually results in higher accuracies [Imp08] and makes it more difficult for forgers to mimic a signature. As digital pens usually implement online-verification [Tol21], we here briefly review online approaches related to digital devices and handwriting as well as digital pens.

Smartphones are plausible candidates for e-signatures. Beside their integrated authentication features, signatures drawn with the finger on the display have been proposed as a feasible alternative [Sae14]. In addition to the visual representation, features such as sound and vibration can also be measured via the embedded motion sensors on the input device [Wan21; Wei21]. Several classification methods have been applied on pictorial and force-related features including Dynamic Time Warping (DTW) [Fan05; Par90; Wir95], Hidden Markov Models (HMM) [Fie07; Jus01; Mur03], Neural Networks [Ala20; Ira13; Lai17; Soe21] as well as Support Vector Machines (SVM) [Bat20; Gru09; Gue15] and Random Forests (RF) [Mee20]. More recently, smartwatches gained attention for authentication purposes. For instance, drawing gestures [Lew16] or signatures in the air [Bai11; Bur18b; Gue21; Lee22; Li21; Moo17; Ngo15; Sun16] can offer an unobtrusive authentication method [Bur18a]. The motion of a pen can also be tracked by wrist-worn devices for authentication [Gri17; Li20] and for recognizing text [Ram21] or gestures [Bec19]. In addition, other smartwatch measurements, such as blood volume changes, can increase the reliability of a signature belonging to a certain person [Rah22]. The bending of fingers contains user-specific features [Ach21] as well as arm movements [Raj16], hand gestures [Yu20] and even simple button presses can differentiate users [Poh15a]. In the vicinity of the pen, the scratching sound that the pen tip makes when it moves across the paper can serve as a distinguishing feature [Din19; Li04; Zha21b]. Via inaudible sound signals from a smartphone, vertical movement variations of the pen can be estimated from the reflected sounds [Che20]. However, it should be mentioned that audio alone is not secured against replay attacks [Wei21] and sensor data alone are not sufficient for a valid signature according to the legal situation [dej22].

Digital pens emit ink on the paper for legally valid contracts and measure related handwriting features, mostly related to the motion of the pen [Gri13; Hsu15; Pan18; Wan12]. Companies like Anoto [AB22], Wacom [Com22], and Stabilo [Gmb22] offer digital pens for education and industrial applications. In addition, 3D Systems introduced a haptic pen input device that can be utilized to write signatures in the air [De 21]. Modular concepts for digital pens have been proposed to allow users to create their own applications [Tey17] with different input and output modalities [Kyu08; Lee04; Wan16a; Xu20]. Integrating fingerprint sensors [Suh16] provides additional security mechanisms, e.g. for biometric authentication in pharmaceutical digital audits [AB22]. The pressure on the pen grip is another source for individual handwriting features [Bas08; Bas11; Hoo03; Kur21] and also pressure applied to the surface is a relevant feature [Bas09; Mur07]. Pictorial features can not only be obtained from cameras [Spe10] but also via magnet tracking, which has been used for sketching on the back of the hand [Sch20a] and signatures [Ket12; Ket10b]. Combinations of audio and motion data have been applied for pen-based handwritten digit recognition [Sch18].

In contrast to related work, we explore the applicability of simple and easy to draw symbols for user authentication. Along with predefined writing trajectories, authentication becomes particularly challenging, because forgers do not have to fully learn another person's handwriting. Furthermore, as few samples as possible should be used for training in order not to stress users who are only able to produce simple x-mark signatures. As previously applied for handwritten digit recognition [Sch18], we combine audio and motion of a digital pen for our approach.

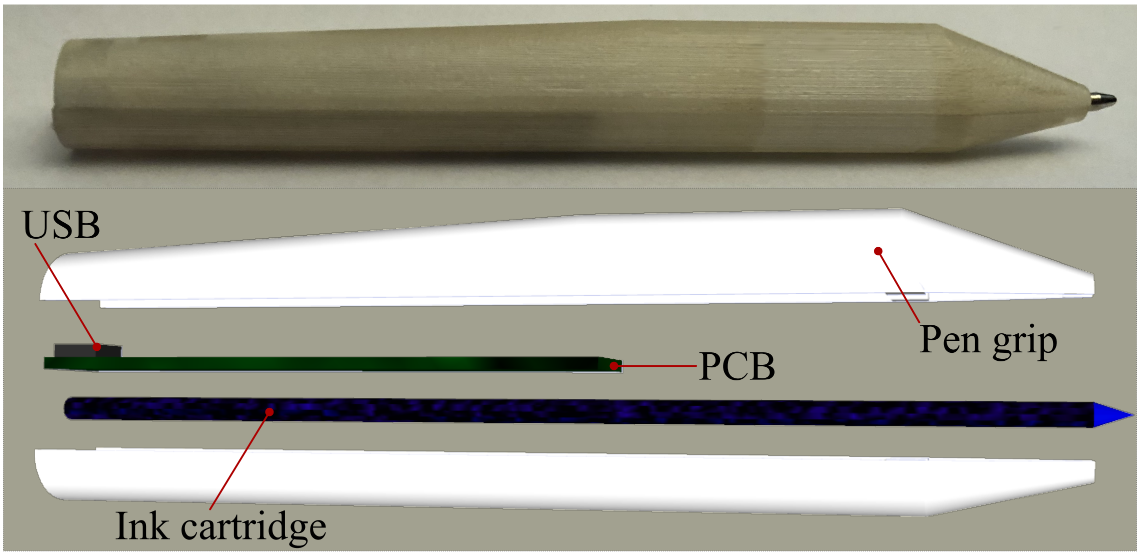

Figure 4.1: The 3D-printed pen prototype (w = 12 $cm \times d$ = 1.5 cm) based on [Sch18] with improved case (added a pen grip).

4.3 Prototype

Our prototype is based on Pentelligence by Schrapel et al. [Sch18]. Data is transmitted via USB to the PC for further analysis. The integrated contact microphone samples audio data at a rate of 7.1 kHz and 3D accelerometer as well as 3D gyroscope measurements at a rate of 800 Hz. Ambient noise is attenuated by the housing, thus mainly scratching sounds are measured during writing. Symbols and strokes are separated by detecting when the pen tip touches the writing surface. For this purpose, a pin is soldered on the spring, which is connected to electrical ground on the PCB. When the pen tip touches a surface, the pin on the spring of the ink cartridge closes a contact to the PCB for identifying writing.

In contrast to the earlier prototype of Schrapel et al. [Sch18] we improved the case by adding a pen grip to constrain the ways the pen can be held. This reduces one source of intra-user variability and helps to capture the way the pen is held in the motion data. Figure 4.1 shows the prototype printed with a Keyence AGILISTA and the corresponding 3D model with the position of the internal PCB.

4.4 User Study

We conducted a study to collect a dataset of various easy to draw symbols. In order to adhere to a skilled forgery setup, we gave instructions regarding the order and direction of the strokes of each gesture. Figure 4.2 on the right, shows our gesture set with predefined writing instructions. Each of our selected symbols just consists of a few strokes and taps, since we aimed to analyze the individuality of symbols that are comparable in complexity to x-mark signatures. For x-marks we use the most frequently occurring writing trajectory in signing tasks [Mar18]. In addition, common gestures from related work [Har08b; Hsu15; Kel06; Lon00] were selected to analyze their applicability for writer identification. In total 19 gestures were selected for our study.

We invited 30 volunteers aged 18 to 56 years ($M = 25.6$ years, $SD = 10.3$ years) including 5 female and 25 male individuals. Two of the male participants were left-handed writers. All participants reported feeling well and rested, which is important because fatigue affects handwriting [Asa13]. Our study procedure was adapted from Schrapel et al. [Sch18]. The study was conducted in a quiet office environment to avoid interfering the internal audio measurements and to allow the participants to focus on their tasks. We began by introducing our prototype and the study procedure. We stated that we intended to collect and analyze gestures with the digital pen. The participants were instructed to repeatedly write a symbol on a 21.0×29.7 cm white squared paper with a box size of 11×14 mm and 15 boxes per row. The symbol and its writing instruction was displayed on a nearby screen. To simplify the selection of individual symbols, the writing of a symbol was confirmed with a short beep and the current count was displayed with a green background. Symbols were cropped in time from the data stream by storing the data from 20 ms before to 20 ms after the pen tip touched the paper. Between two strokes more than 500 ms had to elapse to be recognized as two different symbols. Figure 4.2, left, shows a participant during the study procedure. All 19 symbols were collected in random order. After writing a symbol 20 times, another symbol was displayed until all symbols were written. When a participant accidentally wrote a symbol before hearing the soft beep sound, used the wrong writing instruction or wrote a wrong symbol, the samples were discarded and repeated. This ensured that a clean and balanced data set was created.

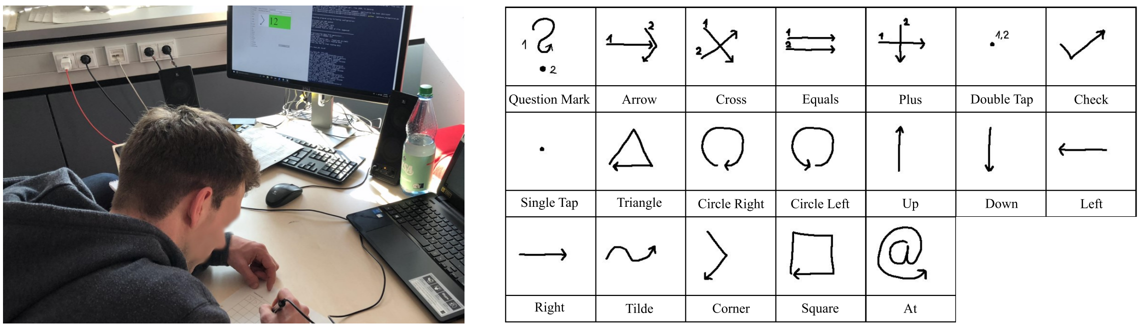

Figure 4.2: Study setup and gestures. The photo shows a participant writing symbols on squared paper. The screen displayed symbols with writing instructions. The set of symbols is visualized on the right. The arrows and numbers represent the writing direction and stroke sequence, respectively.

4.5 Results

We analyzed the collected dataset in order to investigate the individuality of the handwritten symbols. Here, we focus on the use of a small number of training samples with SVM classifiers, as it would be inconvenient, e.g. in the case of signing contracts, to give one's own (symbol-) signature many times before an identification system can be used. Our total dataset consists of 11,400 samples including 380 samples per user with 20 samples for each of the 19 symbols.

4.5.1 Feature Creation

For our analysis, we decided to evaluate different preprocessing steps in the frequency domain for keeping the length of the resulting feature vector constant among the cropped samples [Lam89; Yan09]. To create the feature vector, we first convert the audio signal of each sample into spectral components using a Fast Fourier Transformation (FFT). This is done by splitting the signal of the symbol into one, two, or three equal parts and then applying the FFT to each part. By splitting the data stream, a time component is integrated into the feature vector which may contains additional information about the time depended writing speed and scratching sounds. After normalizing each resulting set of amplitudes by the maximum value and adding the array to the feature vector, statistics are calculated from the raw audio signal. We use the mean and median value as well as standard deviation, and variance of the signal, normalized in the interval [0,1]. In addition, we count the number at which the normalized signal reaches the maximum and minimum value. The time interval between two peaks is also counted. From the counted numbers we calculate the previously applied statistical metrics, normalized by the signal length. This integrates general features of a handwriting into the feature vectors. Then, the same procedure is repeated for the motion data by first summing up the accelerometer axes and gyroscope axes separately. For each summed sensor signal FFTs and statistical features are calculated using the same methods as for

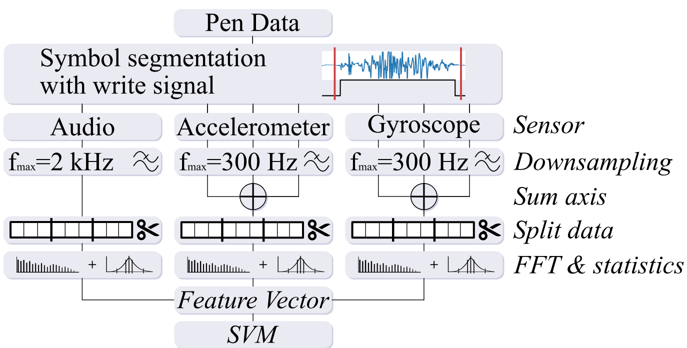

Figure 4.3: The block diagram depicts the algorithm for calculating a feature vector. Each sensor data is processed separately and then concatenated to a feature vector.

audios. The most relevant feature components for audio in handwriting recognition are located below 1 kHz [Sch18]. To correctly measure the relevant features, the signal must be sampled with at least twice this cutoff frequency [Sha49; Whi15]. Thus, we also added the downsampled audio signal to 2 kHz to the analyzed feature set. For motion we downsampled the signal to 300 Hz. Downsampling attenuates high-frequency noise and can may support a classifier to focus on user-specific features. Figure 4.3 visualizes the process of generating the feature vector sets.

4.5.2 Feature Analysis

To analyze the symbols on their suitability for writer identification, we first optimize the SVM parameters and then continue by cross validating on each given feature vector set. The SVM hyperparameters *kernel* (linear or rbf) and cost C (1, 10, 100, 100) are optimized via grid search [Ped11]. The goal of this initial grid search process is to recognize the 30 writers on all symbols as accurately as possible for each previously calculated feature set. Using the established SVM parameters on a given preprocessed dataset, we randomly split the data of each writer into 70 % for training and 30 % for testing. Hence, for each training iteration the training set consists of 266 samples per writer. The SVM is then trained using in total randomly picked 7980 samples to identify the 30 users. We repeat the training and test procedure for 25 times on each corresponding feature set and note down the resulting metrics, i.e. accuracy, precision, recall, and F1-score. As proposed by Levy et al. [Lev18], for each feature set we use the same set of random seeds to obtain comparable results. We performed the procedure on 41 different feature sets resulting in a total of 1,025 tests (41 feature sets × 25 training and testing).

Figure 4.4 on the right shows the results of our feature vector tests sorted by the maximum average F1-score. We obtained an average (macro) F1-score of $F1 = 0.78$ ($SD = 0.05$) with comparable recall ($R = 0.78; SD = 0.12$) and precision

Configuration Lookup

	Audio			*Motion*		
Config	**FFT's**	**Statistics**	**2 kHz**	**FFT's**	**Statistics**	**300 Hz**
1	3	yes	no	1	yes	yes
2	2	yes	no	1	yes	yes
3	3	yes	yes	1	yes	no
4	1	yes	no	1	yes	yes
5	3	yes	no	2	yes	yes
6	2	yes	yes	1	yes	no
7	3	yes	no	1	yes	no
8	2	yes	no	2	yes	yes
9	2	yes	no	1	yes	no
10	3	yes	no	3	yes	yes
11	3	no	no	1	no	yes
12	1	yes	no	1	yes	no
13	1	yes	no	2	yes	yes
14	2	no	no	1	no	yes
15	1	yes	yes	1	yes	no
16	2	yes	no	3	yes	yes
17	3	no	yes	1	no	no
18	1	yes	no	3	yes	yes
19	3	no	no	2	no	yes
20	3	no	no	1	no	no
21	2	no	yes	1	no	no
22	3	no	yes	3	no	yes
23	2	no	no	1	no	no
24	2	no	no	2	no	yes
25	2	no	no	3	no	yes
26	3	yes	no			
27	no	yes	no	no	yes 2x	yes
28	3	no	no			
29	2	yes	no			
30				2	yes	no
31				3	yes	no
32				1	yes	no
33	no	yes	no	no	yes 3x	yes
34	no	yes	no	no	yes	no
35	no	yes	no	no	yes	yes
36	2	no	no			
37	1	yes	no			
38				no	yes	no
39				2	no	no
40				3	no	no
41	no	yes	no			

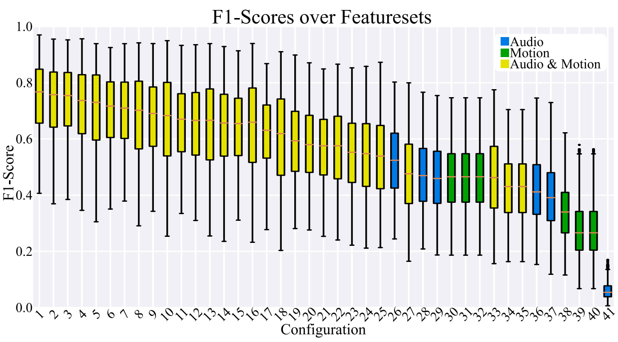

Figure 4.4: F1-score analysis of different feature sets. The tables on the left show the analyzed feature vector generation methods for the visualization on the right. The box plots visualize the F1-scores of different feature vectors including signal splitting, FFTs, statistical features, and downsampling.

($P = 0.79; SD = 0.12$) on the best performing feature set. Stating the average AUC score of $AUC = 0.96$ ($SD = 0.03$) and equal error rate $EER = 0.09$ ($SD = 0.04$) is misleadingly high in our case, because for each of the 30 classes (writers) there exist 114 positive and 3,306 negative test samples. Thus, we will continue reporting F1-scores. In general we derive from the tests that the statistical features have a substantial contribution to the distinguishability of users, and by splitting the audio signal into three parts further improvements are achieved. Combining audio and motion measurements from pens for user identification ($\overline{F1} = 0.74; SD = 0.12$) clearly outperforms audio alone ($\overline{F1} = 0.53; SD = 0.12$) and motion alone ($\overline{F1} = 0.46; SD = 0.11$). On all our 41 SVM hyperparameter optimizations, linear kernels showed better results than Gaussian kernels (*rbf*). In 31 cases a cost parameter of $C = 10$ was selected, followed by five cases using $C = 1000$ (selected by the SVMs trained exclusively on statistical features) and three cases using $C = 100$. A larger C value corresponds to a smaller margin hyperplane and defines how strongly a sample is penalized inside the margin. The best performance is achieved when the raw audio of the samples is splitted into three equal parts where of each statistical features and FFTs are added to the feature vector. Data of accelerometer axes and gyroscope axes are downsampled to 300 Hz and summed up respectively. The FFT and statistical metrics are calculated from each of the two summed signals and added to the feature vector. Splitting of the summed motion signals is not applied. The grid search on this best performing feature set resulted in using a *linear kernel* with a *cost* $C = 1$ for the trained SVMs. For the best performing feature set we repeated the grid search test on three, four, and five-degree polynomial kernels and a sigmoid kernel, again obtaining the best results with a linear kernel with a $C = 1$.

4.5.3 Symbol Analysis

We now analyze symbols with respect to their suitability for writer identification by using the best-performing feature set and corresponding SVM hyperparameters. The cross validation test is repeated by using one, three, five, and ten training samples. For each training size on each symbol we perform 25 tests by using a fixed set of random seeds to pick training and test samples. In total 1,900 multi-class SVMs were trained and tested (4 training sizes × 19 symbols × 25 tests). We analyze the influence of the training set size on the F1-score of the classifiers for each symbol.

In Figure 4.5 on the right, the symbols are displayed sorted by the highest average F1-score with one training sample. On average, we obtain an F1-score of $F1 = 0.55$ ($SD = 0.065$). *Arrow* achieves the highest F1-score ($F1 = 0.66; SD = 0.035$), followed by *Circle Left* ($F1 = 0.658; SD = 0.035$). *Cross* showed an average performance within the overall confidence interval ($F1 = 0.57; SD = 0.029$). *Right* had the worst performance ($F1 = 0.47; SD = 0.035$). A one-way ANOVA revealed a statistically significant difference of the F1-scores between at least two symbols ($F(19{,}25) = 61.02, p << 0.001$). Hence, we performed a post-hoc Tukey test to identify which symbol pairs have a significantly different F1-score. *Cross* shows a significantly different F1-score to twelve of the symbols including: *Arrow, At, Circle Left, Double Tap, Down, Equals, Plus, Question Mark, Right, Single Tap, Square, Triangle*. The symbol with the best performance, *Arrow*, showed only no statistically significant difference to *Circle Left* ($p = 0.9 >> 0.05$).

Figure 4.5 on the left, depicts the impact of the number of training samples on the F1-score. With increasing training sample size we obtain more accurate results. A one-way ANOVA revealed a statistically significant difference of the F1-scores between at least two training sizes ($F(4{,}475) = 3940.85, p < 0.001$). A subsequent post-hoc Tukey test showed a statistical significant difference of the F1-scores between all group pairs meaning that a larger training set significantly improved the classifier's performance, as one would expect. In addition, we wanted to find out whether the precision and recall measures have comparable performances. A one-way ANOVA indicated statistically significant differences between both measures ($F(8{,}475) = 3125.71, p < 0.001$). A subsequent post-hoc Tukey test found no significant difference between recall and precision when ten training samples are used ($p = 0.1 > 0.05$). On average the precision is slightly higher by $\Delta = 1.2$ % ($SD = 0.38$ %), which still indicates an acceptable tradeoff between the two measures. Ideally, the classifier would reject all other writers and only accept true samples of an individual. A higher precision means that the SVM is more accurately rejecting other writers. While a higher recall results in a more accurate identification of the writer. Therefore, precision is more important in authentication tasks. The F1-score is the harmonic mean of both measures and reaches in an ideal case a value of 1.

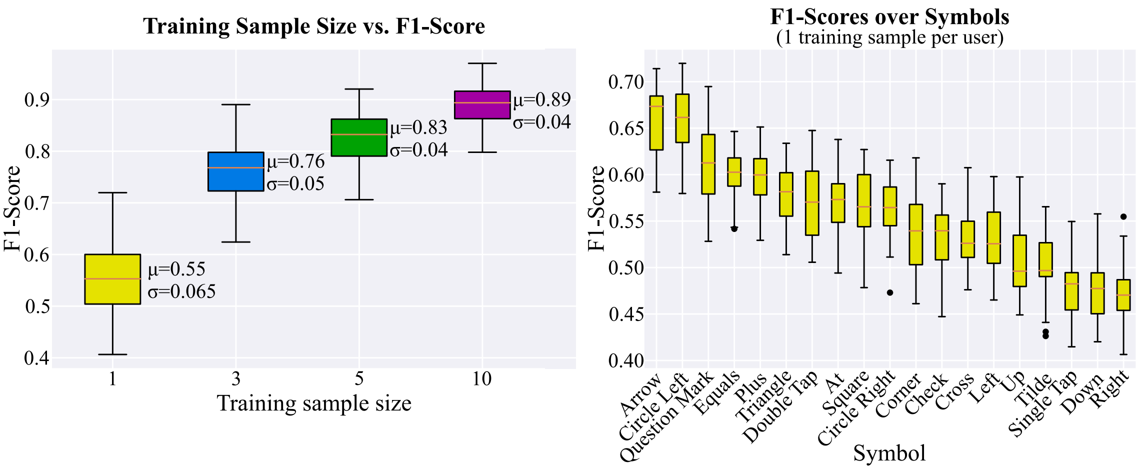

Figure 4.5: Training size and symbol analysis. The left plot depicts the impact of training sample count on the average F1-score over all symbols. The right plot compares the F1-scores of each symbol with one training sample. We use the best performing SVM hyperparameters and feature set for the results. One training example was chosen to illustrate the most challenging use case.

4.5.4 Majority Voting

As contracts might require a person to sign a document several times, we analyze the impact of averaging classification results. For this purpose, we used the predicted results of the previous tests and average all possible combinations without duplicates by the formula: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Where n is the number of predictions and k is the number of voting samples. Using each previous training sample sizes $t = [1,3,5,10]$ and the corresponding test sample sizes $n = [19,17,15,10]$, we analyzed $k = [1,3,5,7,9]$ voting test samples. The user is then determined in a voting result based on the maximum occurring class.

In Figure 4.6, left, the impact of majority voting on the F1-score and training sample size is plotted. The lines are the determined mean values of the F1-scores for each training sample size. The colored shades indicate the standard deviation around the mean values. For each training size a one-way ANOVA revealed significant differences between at least two groups of majority voting F1-scores ($F(5{,}475) = [727.5, 1805.9, 2405.0, 2263.5], p < 0.001$). A subsequent Tukey test found significant differences for each training size between all corresponding numbers of voting samples. There was no significant difference between 7 and 9 voting samples using 10 training samples ($p = 0.127$). However, this result can be related to the limited amount of 20 samples per symbol and user. To exemplify the impact of majority voting, Figure 4.6, right, depicts voting results on the symbols *Arrow*, *Cross* and *Right*. On average three voting samples increase the F1-score by $\Delta = 8.9\ \%(SD = 3.3\ \%)$. We derive that majority voting positively effects writer identification and that signing a contract with three symbols allows a more reliable user identification than with just one symbol. Symbols such as *Arrow* were found to have a significantly higher potential of accurately identifying users than *Cross* with our approach.

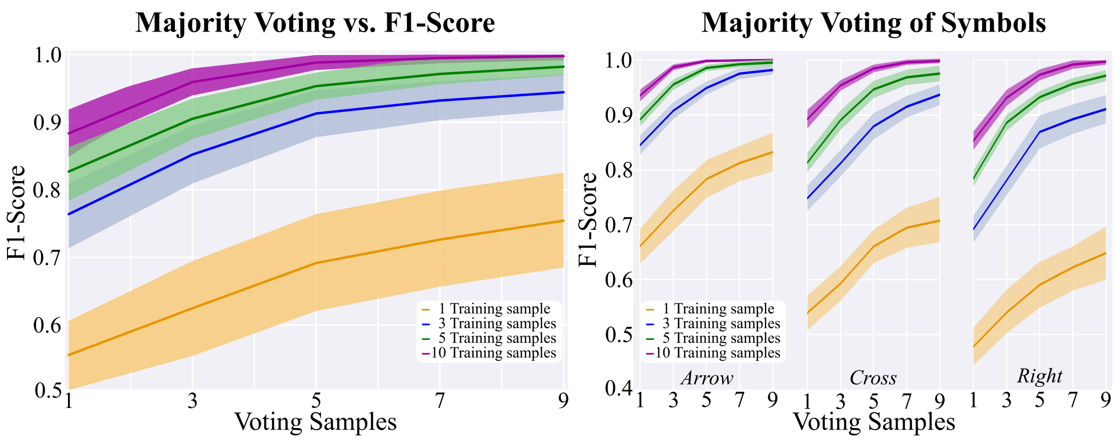

Figure 4.6: Majority voting analysis. On the left, the average impact of the number of voting samples on the F1-score across all symbols. The plots the right depict the impact of the number of voting samples on the symbols *Arrow*, *Cross* and *Right*.

4.6 Discussion

Identifying writers by handwritten symbols is a major challenge in cases of suspected fraud. So far, the focus has been on ordinary signatures, but in cases of illiteracy or disabilities, a writer may be limited to signing with x-marks or similar easy to draw symbols. At least three people have to be involved in the process of a valid x-mark signature [dej22]. This includes the writer, the contractor, and a witness. Cases of suspected fraud are associated with a high financial expense, as the signatures have to be analyzed using forensic methods and the relationships of witnesses have to be investigated. Additional security mechanisms can contribute to identify forgeries. One challenge in collecting samples of handwritten symbols is that writers of x-mark signatures may tire faster. Thus, a classifier must be able to reliably distinguish writers with as few samples as possible.

Our results show that combining audio and motion data of digital pens can more reliably identify writers by handwritten symbols than single sensors. This confirms the finding by Schrapel et al. [Sch18] that both sensors contain relevant and specific features of a symbol signature. In addition, this is in agreement with the results of Muramatsu et al. that combined sensor approaches increase the reliability of online verification approaches [Mur07]. Splitting the incoming data stream is advantageous for scratching sounds in order to integrate a time component into the features. In contrast, higher recognition rates were achieved when motion data of a symbol were not splitted. In addition, downsampling the measurements to 300 Hz was beneficial on motion data, while using the raw audio signal at 7.1 kHz resulted in higher scores. Compared to previous approaches that used spectral analysis [Lam89; Yan09], our results show the importance of the time component in signature verification tasks. Additional statistical features of both sensors can further increase the accuracy of the classifier. Hence, the time component of the signals is more important for audio measurements, while the high-frequency components of motion data contain less information. This can be attributed to the writing speed. Vibrations on the paper are more likely to occur at lower frequencies (up to 150 Hz). Produced scratching sounds are modulated by the writing speed [Li04]. Small irregularities of the paper surface act as a carrier that influence the pitch of the produced scratching sounds. In addition, the volume is influenced by the pressure on the paper [Hwa12]. Together with loop-back sounds that propagate via the bones of the hand, resonance properties of the writing surface and the way the pen is held, unique features are produced. This allows relatively accurate identification of a writer with just a few training samples.

Our analysis of the individual symbols shows that certain symbols include a higher diversity of individual characteristics for our participants. This can be attributed to the individual writing style and the way of holding the pen. While the x-marks achieved a uniform performance over the analyzed symbols, the F1-scores of the three best performing symbols *Arrow, Circle Left, Question Mark* exhibit a significantly higher reliability in identifying writers. However, arrows on contracts are commonly used to indicate signing fields and question marks may also be misinterpreted. In addition, according to the legal situation in Germany, a contract must be signed with three crosses for validity [dej22]. With three crosses as a training set, the F1-Score already increases to about 75 % in comparison to 55 % with only one training cross. In the case of suspected fraud, it would be required to write at least three crosses again and select a user based on majority voting. With this approach the F1-Score increased to over 80 %. The more training and test samples are given for majority voting, the more reliably the 30 users were identified. With 7 test majority samples, we already reached an F1-score of over 90 % on x-marks by using three training samples. It should be noted that x-mark signatures are particularly challenging due to their structural simplicity. A skilled forger can easily learn the trajectories of a symbol [Hil92; Hub99; Mar18]. However, the scratching sounds that propagate via the loop-back channel over the hand bones contain individual features that make forgeries more challenging.

Compared to previous research on handwritten digit recognition with audio and motion data [Sch18], we showed that combining both sensors allows identifying writers. Since these features are writer specific, generalization of handwriting recognition classifiers is challenging. Schrapel et al. circumvent this problem by retraining their neural networks on samples of the writer. This assumes that the current writer is known and individualized classifiers already exist. However, it could not be determined whether the current writer is already known to the system. Our research has closed this gap and shown that the writer can be detected by individual symbols. This detection allows automatic selection of a user-specific classifier and support for multiple users of a pen. In addition, the presented approach opens up new application scenarios.

We will now discuss the components and challenges of such a possible future x-mark signature verification application. In this context, our prototype may serve as an additional security element in cases of suspected fraud. The system components involved are illustrated in Figure 4.7. A full implementation would have to be integrated into a suitable public-key and document archive infrastructure. When signing a contract the pen would encrypt and cryptographically sign the measured data along with a unique pen hardware identifier and timestamps. The date and time on the physical contract allows assigning a digital timestamp of the pen to a signing event. Furthermore, the metadata of the contract can serve to enable unique identification. To additionally prevent subsequent modification of the stored pen data, each signing event could be irreversibly stored in a blockchain or on a trusted, secured server. Although both a blockchain and a server structure would be possible on the backend [Wüs18], there is a major challenge in making the stored data immutable with a central solution. An attacker could try to delete or modify the data to prevent later verification.

When signing a contract, only the person authorized to sign with crosses uses the digital pen. In case of suspected fraud, all persons involved (e.g. contractor, witnesses and writer) would write at least 3 samples of the symbol, e.g. the x-mark, with the digital pen that was used when signing the contract. These samples are used to verify the signature on the contract. As shown in Figure 4.6, the more samples are written, the more reliably a person can be identified by majority voting. To prevent the writing of invalid samples during this observation, another neutral person (investigator) monitors the process. The neutral person ensures that all writers give samples of the symbol with the same writing instructions as found on the physical contract in the same environment, or at least with comparably low ambient noise level in a similar environment. While the pen is not in use, the ambient noise level may be measured so that comparable conditions can be created later. Likewise, signatures could be automatically rejected in noisy environments. In order to verify a symbol signature, data previously stored in the blockchain or on a trusted server is used to train an SVM. This allows to later determine a probability that the signature on the contract belongs to a specific person in question. The more contracts were previously signed with our digital pen by an individual, the more reliable the identification system becomes. Since audio modulates the scratching sounds on the surface and by propagating though the hand, the signal includes individual characteristics. However, in case of intermediate hand surgeries or diseases those characteristics might change. Thus we do not consider such cases. In addition, it may happen that a writer merely claims not to have signed a document. In these cases, the person would try to fool the system by, for example, holding the pen differently or pressing the pen tip harder on the sheet of paper. It is likely that the resulting x-mark signatures differ from other previous contracts of the writer. Differences could also be determined by forensic analysis. Furthermore, stored signatures may already exist in the backend from other valid contracts that can be used for analysis. Since the handwriting changes in the course of life [Gam80; Sri08] a time limit of used signatures for training an SVM would be necessary. Besides this application scenario, an additional security mechanism could be implemented, e.g. for biometric authentication in pharmaceutical digital audits [AB22]. In addition to the fingerprint sensor [Suh16], our approach would recognize x-marks while filling out forms. Since fingerprint sensors in mobile devices can be easily fooled by fake fingers, e.g., made of silicone [Goi18], our mechanism could provide additional confidence that the pen is held by an authorized user.

The results of our study and the example application scenario show limitations of the presented prototype. First, the acoustic signals might also include sounds from the environment. Even though the scratching sounds are considerably louder than environmental noises while writing, our approach is limited to quiet office environments. The influence of environmental noise has to be further investigated in

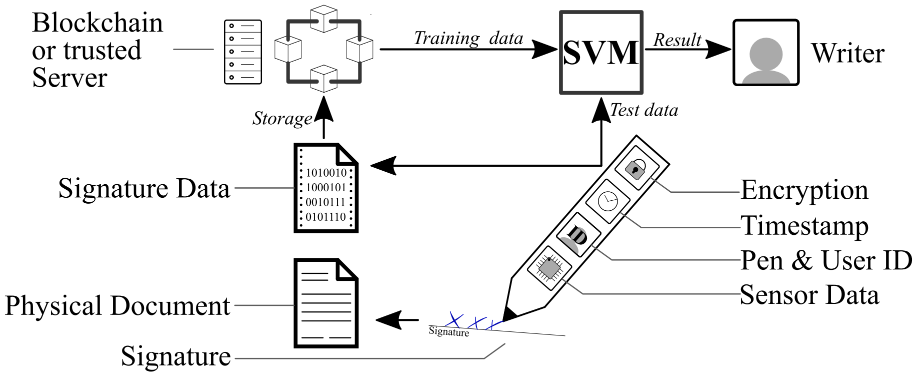

Figure 4.7: Proposed application scenario: The encrypted measured motion and audio data of a signature is stored in the blockchain or on a trusted server and can be retrieved for subsequent writer identifications. The data is encrypted and used for training an SVM. Additional signatures or symbols are then tested on the trained SVM.

future studies. In addition, the paper used for signing a contract can have an influence on the measured scratching sounds. Therefore, scratching sounds on different surfaces and papers should be further investigated. Furthermore, our approach must be limited to small numbers of writers. In our use case, all contract partners are known and therefore a classifier can more easily identify individual features of the writers. We assume that in a usage scenario with a significantly larger number of writers, there may be more false positive classifications. Hence, we limit our approach to contracts which are likely to involve a limited number of contract partners. To further increase the reliability with a significant higher number of users the pen could be extended to measure the force of holding the pen in the hand [Bas12; Hoo04] and a fingerprint sensor [Suh16].

4.7 Conclusion

This chapter presented an approach to identify users via handwritten symbols using a digital pen that measures audio and motion data. A study with 30 participants was conducted to evaluate the approach on 19 candidate symbols. The collected dataset consists of 11,400 samples resulting from 380 samples per user with 20 samples for each of the 19 symbols. The proposed method has three main contributions: (1) a preprocessing algorithm for combining audio and motion data from digital pens in the frequency domain to identify writers via handwritten symbols, (2) an analysis of different feature sets fed into SVMs on their suitability for writer identification, and (3) an examination of various symbols including x-marks on their ability to identify writers based on small training sets in an online-verification task.

We found that combining scratching sounds of a pen tip when moving on a sheet of paper with the motion of a pen outperforms single sensor approaches to identify writers via handwritten symbols. Signal representations in the frequency domain were fed into SVMs to establish our results. To the best of our knowledge, we are the first to unite audio and motion data for writer identification via handwritten symbols. The approach achieves an average F1-Score of 76 % when using only three training samples and 17 test samples per user and symbol. Applying majority voting on three test samples increases the average F1 score to 87 %. We were able to show that certain symbols have more distinctive features despite having similar writing trajectories. The symbol *Arrow* achieved the highest F1-Score of about 91 % with three training samples and applying majority voting on three test samples. X-mark signatures reached significantly lower F1-Scores of about 82 % in the same configuration.

The proposed system may be used as an additional security mechanism for illiterate persons or persons with motor disabilities to support investigators in cases of suspected fraud. In this application scenario, the challenges of writer identification are particularly high. As few symbol samples as possible must suffice to identify the writer. In the ideal case, the signature on the contract should be sufficient to train a classifier. The presented approach allows this, but is limited to low-noise environments, such as offices. Therefore, further research is required in the future on how external noises can be suppressed. A long-term study is also required to investigate the change of handwritten x-mark signatures of single persons as handwriting changes in the course of life.

CHAPTER 5

The Skin as a Drawing Space for Small Displays

The previous two chapters introduced surface-related scratching sounds as complementary features to motion that support the recognition of handwritten information and also identify writers. However, it was not possible to determine the position of the information on the paper resulting in sketching was not supported. This chapter introduces the back of the hand as a suitable input and drawing surface for small displays. For this purpose, a wireless pen prototype is presented which detects touch events with the skin. An embedded magnet is used to determine the relative pen position on the back of the hand to a magnetometer of a wrist worn device. We compare different magnet tracking algorithms and optimize the magnet's field strength in relation a wrist-worn magnetometer. We then test the prototype in a user study where participants had to perform pointing and selection tasks as well as sketching. We then continue by evaluating our implemented software prototype in a usability experiment in order to answer the following research questions.

> Is the back of the hand an appropriate writing surface for pens to extend the interaction space of tiny wrist-worn displays?
>
> Is magnet tracking accurate enough to deploy a drawing surface on the back of the hand?

This chapter is based on the bachelor thesis by Florian Herzog [Her19] as well as Steffen Ryll [Ryl20] and a late breaking work at *CHI 2020* [Sch20a]. In their theses Florian Herzog evaluated different magnets and tracking algorithms and Steffen Ryll improved the computational performance on the smartwatch and implemented and evaluated the sketching application.

- Schrapel, Maximilian & Herzog, Florian & Ryll, Steffen & Rohs, Michael: 'Watch My Painting: The Back of the Hand as a Drawing Space for Smartwatches'. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.* CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, 2020: pp. 1–10

Figure 5.1: Suggested application scenario of the pen prototype PentelliZen: When the pen tip touches the skin, strokes can be drawn. A cursor on the display shows the corresponding position of the pen. To maximize visibility, menus are hidden at the edge of the display. The size of the drawing area, color, and size of the strokes can be adjusted via mid-air interaction with spatial menus.

5.1 Introduction

Modern smartwatches are mainly used in sporting activities, health tracking, and with simple interactions such as answering phone calls, navigation, or music control. Due to their functionality of phone-independent mobile communication they can be used in a variety of applications, but face challenges associated with the small interaction area of the touch display. For example, tasks like text messaging are still difficult, because the finger covers several keys and users cannot see what they are actually typing [Lei15]. As an alternative option to typing, drawing is a very personal way to reply to incoming messages or to create, for example, an individualized digital birthday card [Kie13]. However, drawing requires very precise input, which amplifies the challenges of small displays, e.g., with regard to the *fat finger problem* [Wig07].

Both hardware and software approaches have been proposed to address challenges related to tiny displays. While software solutions are easy to deploy, they are mostly confined to the display space. Additional hardware can extend the interaction space but requires further equipment on the device. Fortunately, modern smartwatches are already equipped with a variety of sensors, which can be utilized for input. Using those existing sensors for precise input opens up new interaction opportunities without reducing the wearing comfort. For example, previous works employed embedded magnetometers for basic interactions. However, the achievable tracking precision and whether magnetometer-based sensing can outperform direct finger interaction were up to now unclear.

We present *PentelliZen*, a novel digital pen for precise cursor input on smartwatches. The back of the hand is used as an interaction space to avoid occlusion of the small display. By integrating a magnet and motion tracking inside the pen, an embedded magnetometer on the watch can be used to calculate a relative cursor position on the display. Due to the form factor of a conventional pen and the avoidance of additional equipment attached to the wrist the wearing comfort is not reduced. To prevent unintended input, a touch sensor, connected to the gold-plated pen tip, detects skin contact and distinguishes drawing on the back of the hand from mid-air gestures. The pen sends its data directly to the watch via Bluetooth. Special attention has been paid to design the pen so that the skin contact of the pen tip on the back of the hand is comfortable. An example use case is shown in Figure 5.1.

In this chapter we investigate drawing as an example application, because drawing and sketching place particularly high demands on input accuracy: Pens are a natural input device for this purpose. Drawing as a use case illustrates the potential of the approach. We evaluate different magnetic and algorithmic parameters to optimize the tracking performance. Additionally, drawing interfaces need to provide a broad functional variety of drawing tools, which increases the complexity compared to other tasks. To facilitate the selection of drawing tools and the setting of parameters, our prototype allows for interacting in three-dimensional space above the back of the hand in addition to drawing with skin contact. We compare the performance of the proposed approach to direct finger input in a user study with twelve participants. A subsequent qualitative usability study with five participants serves to evaluate the implemented interaction approach. We show that the back of the hand is a natural interaction space similar to a touch pad.

5.2 Related Work

Methods for counteracting the fat finger problem on tiny displays can be divided into software and hardware approaches. By extending the software of the device, simple interaction techniques can be realized that do not require further equipment. For instance, enlarging relevant parts of input areas is a simple but effective method for text entry on ultra-small displays [One13]. Especially with a large number of icons or emojis, zooming into areas of interest is faster and more precise than swiping [Poh16]. For text entry, grouping letters increases the input rate, as the impact of the fat finger problem is reduced [Che14; Dun14; Hon15; Qin18]. Recognizing which finger touches the screen can reduce the number of interaction elements [Gil17; Par20]. Furthermore, tap and gesture typing with statistical decoding can be used to automatically correct input errors [Gor16]. Display occlusion can be counteracted by showing the hidden information under the finger on the remaining display space [Lei15]. Through cursor entry [Shi16], back-of-device interaction [Bau09] or wrist motion tracking [Gon18], interaction elements can be kept visible. Moreover Zhang et al. proposed a technique to use motion and audio signals from smartwatches to recognize gestures [Zha16a] around the device.

Additional hardware on the finger segments can enable extended opportunities for interacting with wrist-worn devices [Won17] or to track fingers in virtual reality [Lyo20] and for haptic actuation [Pec17]. Rings can be utilized for scrolling activities and pointing tasks [Ash11; Che13; Chu18; Har09; Ket10a; McI19; Par19a; Par20; Rey18; Tsa16b; Zha17; Zha16c], gesture recognition [Che18; Che19b; Chu18; Dar15; Gup18; Ket13; Ket10c; Oga12; Tsa16a; Yoo16b], finger identification [Gup16; Par20; Par19b], or to enhance virtual keyboards [Abd20]. Hardware attached to the fingernails can be used as a touchpad [Kao15], for near-surface tactile feedback [Wei11], finger tracking [Che16], or even to enable precise input via a miniature stylus pen [Xia15]. Touch functionality can also be integrated into clothing, e.g., for interactions during sporting activities [Sch16]. Additional tangible widgets [Hwa13a] can enable new interaction techniques and be applied for authentication around mobile devices. Integrating sensors into the watch wristband is a promising approach as it can expand the input and output space [Gon16; Per13; Sav17; Str15; Zho16]. Similarly, the bezel of a watch can provide natural interaction techniques [But08; Dar16; Oak14; Yi17], where embedded systems also enable touch events on the back of the hand [Kni14; Oga15; Sri17]. Furthermore, a movable mechanical display can offer various opportunities by using pan, twist, tilt, and click [Xia14]. By integrating microphone arrays into devices, Nandakumar et al. have shown that fingers could be tracked around the device [Nan16].

Several of the mentioned techniques use magnetic field sensing [Abd20; Ash11; Che13; Che16; Che18; Che19b; Chu18; Dar15; Hwa13a; Ket13; Ket10a; Lyo20; McI19; Par19a; Rey18; Wei11; Yoo16b; Yuy15], which has also been proposed for stylus pens [Abe16; Hwa13b; Yoo16a] or head posture [Han20] and tongue tracking [Seb19]. It remains challenging to optimize the magnet tracking performance, since different influences impact tracking accuracy. For example, the dimensions and physical properties of a magnet are an important limiting factor [Cam13] as well as sensor noise and the number of magnetometers used for tracking [Fan16]. However, it has not yet been clarified whether proposed algorithms [Abe16; Cam13; Fan16] are sufficient for precise interactions, such as drawing on the back of the hand. There has been research on using cameras and motion sensors for precise drawings on paper [Wu17], yet these are difficult to miniaturize for use with wearables. Even without position tracking, digital pens can recognize handwriting by using audio and motion data [Sch18].

To summarize, considerable effort has been made to investigate interaction techniques for wearables, in particular using magnet tracking. However, up to now the focus has been on coarse input tasks or limited around-device gestures, rather than on addressing the requirements of precise interactions. Moreover, a detailed performance comparison of around-device magnet tracking algorithms is still missing. We close this gap and propose new interactions based on cursor input using magnet tracking on the back of the hand with our pen. Furthermore, a usability study is conducted to determine the applicability in two- and three-dimensional space for back of the hand interactions with smartwatches and our pen prototype *PentelliZen* for drawing tasks.

5.3 Position Tracking

Since drawing requires a particularly high tracking precision, we evaluate known algorithms for magnetic tracking.

5.3.1 Algorithm Selection

To identify the best tracking algorithm, three different approaches from related work [Abe16; Cam13; Yoo16a] were implemented on a *Huawei Watch 2*. In the selection process for the most suitable tracking algorithm, the intended use with a stylus pen was especially considered. The three candidate algorithms we considered were the ones by Abe et al. [Abe16], by Camacho and Sosa [Cam13], and by Yoon et al. [Yoo16a].

We first chose the 2D tracking algorithm by Abe et al. [Abe16], assuming that the back of the hand approximates a flat plane. When all influences of terrestrial magnetism are constant, the magnetic moment can be eliminated using Coulomb's law, and a root-finding problem remains. In contrast to Abe et al., who apply the bisection method, we use Newton's method on all algorithms, as it requires less computational power, which is advantageous for mobile devices. While Abe et al. only consider the height of the magnet [Abe16] Camacho and Sosa also use its shape in their equations [Cam13]. Since cylindrical shapes can easily be embedded in stylus pens, this shape and approach was selected.

Furthermore, we selected the 3D tracking algorithm by Yoon et al. [Yoo16a] which was used in various projects [Fan16; McI19]. In contrast to the other chosen methods, this approach takes into account the magnetic moment, which determines the orientation of the magnet. However, the orientation of the magnet in three dimensions is required, which Yoon et al. calculate using an additional magnetometer [Yoo16a] and Fan et al. [Fan16] as well as McIntosh et al. [McI19] by placing several magnetometers around the interaction surface.

5.3.2 Algorithm Evaluation

To examine the optimal magnetic parameters and to evaluate the performance of the selected algorithms with the magnetometer embedded in the watch, a tracking experiment was conducted. Based on the dimensions chosen by Abe et al. [Abe16], cylindrical magnets with a diameter of 8 mm and heights of 16, 20, and 32 mm with a fixed magnetization degree of N48 were selected. Since a higher magnetization results in a higher magnetic flux density, a magnet with a diameter of 7.5 mm and a length of 20 mm with a magnetization of N52 was also used in the experiments. The experimental setup is shown in Figure 5.2.

The first experiment was intended to measure the accuracy of the algorithms in 2D space. The *Huawei Watch 2* was placed in a planar coordinate system with the embedded 9-DOF smartwatch IMU in the origin. Within the proposed interaction area the cylindrical permanent magnet (diameter 8 mm, height 20 mm) was moved in 1 cm steps within a grid of 4 to 12 cm in x-direction and -5 to 1 cm in y-direction. The grid size was selected according to the average of anthropometric measurements of the back of the hand [Aer95]. At each position, with the magnet's north pole towards the ground, the calculated distance was logged for all three algorithms. Deviations, resulting from inaccurate magnet orientation tracking for the 3D tracking algorithm were eliminated due to the chosen test setting.

A second experiment was conducted to investigate the optimal magnetic parameters. The previous test was repeated for different magnets with the 3D tracking algorithm. Based on the results, the two most promising magnets were evaluated again in an tilt angle of 45°, which corresponds to a more realistic usage for stylus pens. Based on the final selection, the influence of the magnet's offset in z-direction was then investigated by measuring the positions again with an offset of 5, 10, 15, and 20 mm on the z-axis.

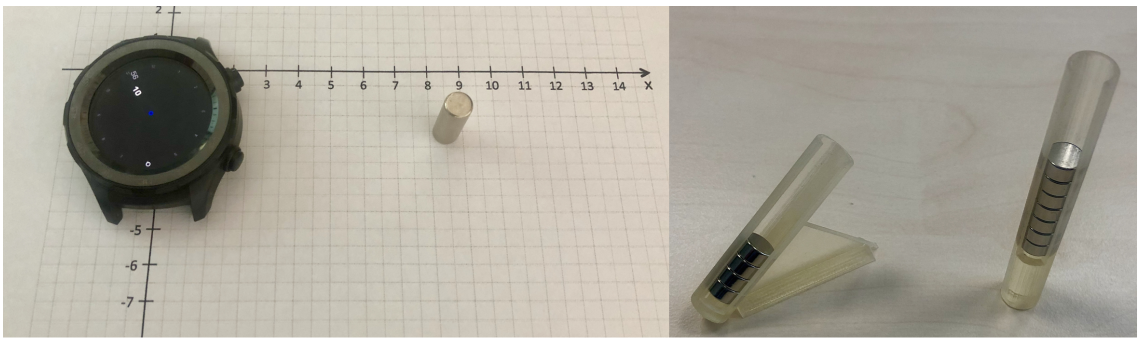

Figure 5.2: Experiment setup: The left picture shows the coordinate system used in the experiments. The embedded magnetometer of the Huawei Watch 2 is placed in the origin and the magnet is moved in 1 cm steps in x and y direction. The right image shows the printed cases for a 45° inclination of the stacked magnets and with a specific z-offset.

5.3.3 Tracking Results

Three dimensional tracking requires including the magnetic moment in the equations. This results in better tracking performance compared to the other approaches, as shown in Figure 5.3a. A Bonferroni-Dunn post-hoc test confirmed a significant difference for the 3D algorithm. Although the inclusion of the magnet shape in the equations causes smaller fluctuations, there was no significant difference between the two 2D approaches $F_{1,26} = 0.312$. The experiment shows that the 3D position estimation has superior performance even on planar surfaces. Hence, the experiments are continued with the approach by Yoon et al. [Yoo16a].

Figure 5.3b visualizes the tracking performance in relation to the used magnet. A higher degree of magnetization causes sensor saturation in the proximity of the measuring unit. As a result the magnet [8×32 mm; N48] achieves a significantly worse tracking performance according to a Bonferroni-Dunn post-hoc test. On the other hand, when the degree of magnetization is too low, the applied 3D algorithm on the magnet [8×16 mm; N48] cannot find a solution over the full tracking area. The maximum distance to the sensor in x-direction is 5 cm, which does not cover the average of anthropometric measurements of the back of the hand [Aer95]. Therefore, a sweet spot must be found that has low and constant distance error rates within the interaction area. The two remaining magnets ([7.5×20 mm; N52],[8×20 mm; N48]) do not differ significantly in their performance [$F_{1,26} = 0.72$, ns] and are examined in the subsequent test.

Previous tests focused on ideal planar surfaces, which is not appropriate in our back-of-the-hand scenario. By tilting the magnet, its field lines are measured on all sensor axes. The results are shown in Figure 5.3c. The solid magnet [7.5×20 mm; N52] can achieve adequate performance within the full interaction area. When 4 mm magnets are stacked, the accumulated force gain declines with the number of magnets. As a result the stacked magnet [8×20 mm; N48] achieves a lower magnetic force. In inclined position the signal-to-noise ratio of measurements decreases, which leads to a maximum tracking distance of 4 cm to the sensor. Hence, the solid magnet [7.5×20 mm; N52] is selected for the final experiment.

The back of the hand does not form an ideal planar surface. Therefore the influence of the z-position on the computed 2D position must be taken into account. As shown in Figure 5.3d, an increasing relative error can be observed with increasing z-offset. A z-offset starting from 10 mm shows a significant performance decrease according to a Bonferroni-Dunn post-hoc test. This is caused by a decreasing magnetic flux density. However, the approach still achieves an acceptable maximum error of 6 % up to a 15 mm offset.

The results of the experiments show that the 3D tracking approach outperforms the 2D approaches and that the tracking performance depends on an appropriate magnetic field strength. These influences can lead to a smaller tracking distance in

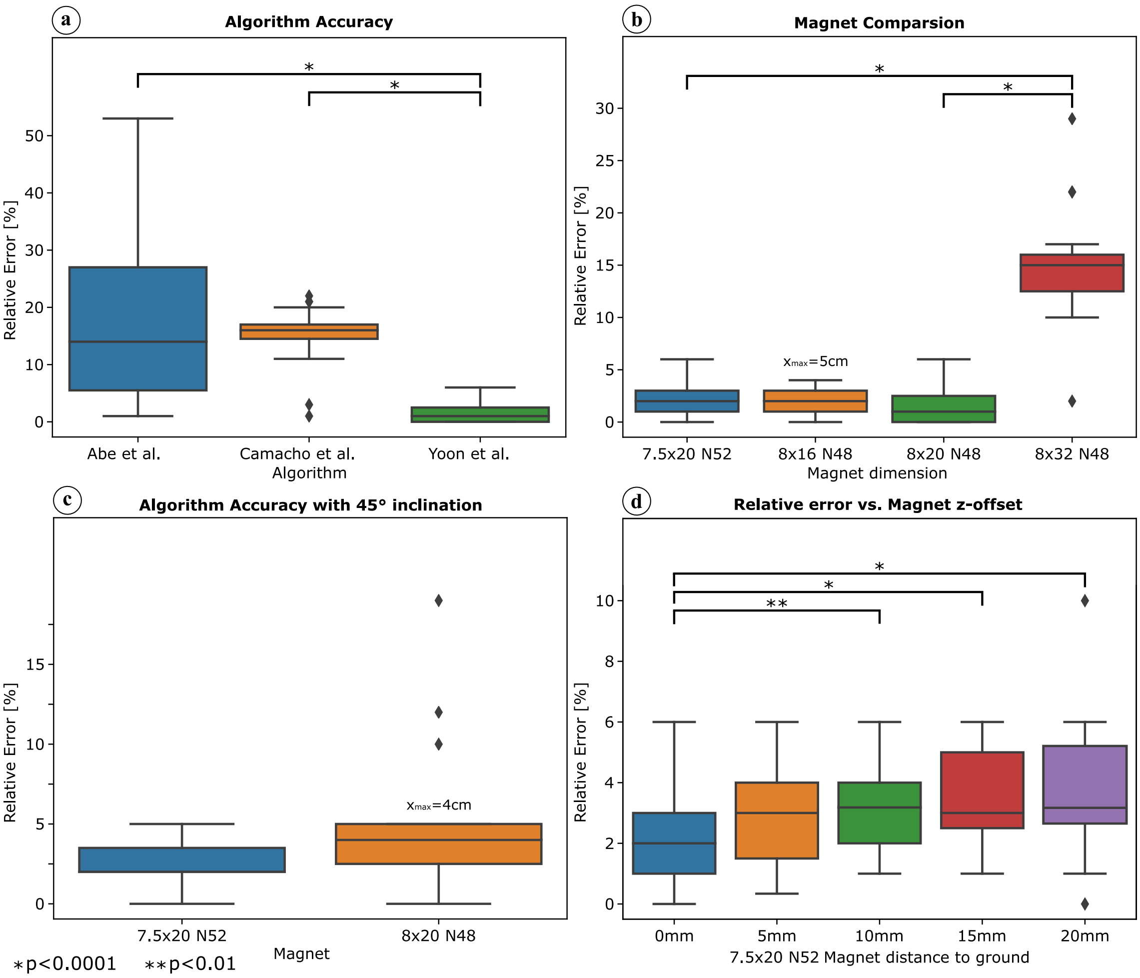

Figure 5.3: Magnet evaluation in relation to the relative distance error, calculated by dividing the absolute calculated distance by the real position: (a) shows a comparison of different algorithms, (b) examines the performance in relation to the magnetic field strength, (c) repeats the test in inclined magnet orientation, and (d) examines the influence of the z-offset on the calculated distance.

an inclined orientation or with a non-zero z-offset. The magnetic field strength has to be adjusted according to the magnetometer's sensitivity. Improved sensor technology might allow the use of smaller magnets or of a wider range of magnetic field strengths. The performed experimental protocol can be applied to future developments to maximize the tracking performance in a specific use case and with specific hardware. Figure 5.4 visualizes the distribution of the relative error on the back of the hand by using the algorithm of Yoon et al. [Yoo16a] and the final selected magnet [7.5×20 mm; N52]. An acceptable tracking accuracy is achieved, which is to be further verified in the subsequent user study. However, a considerably less precise position estimation is achieved at the knuckle level. A tracking error of 6 % at a distance of 9 cm corresponds to a deviation of 5mm. There is also no planar surface, which is still sufficient to implement an interaction such as scrolling at the level of the knuckles. Thus, the proposed interaction technique on the back of the hand can be compared to a touchpad on notebooks, where some models implement a scrolling area on the right edge. However, it can already be stated that the presented technique is limited to the near field of the watch and the initial idea of a tracking system for pen and paper is not possible with magnet tracking.

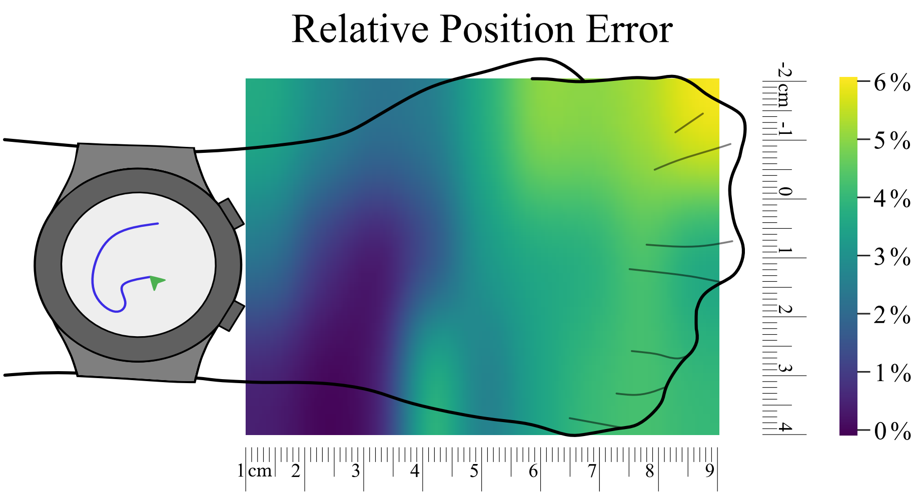

Figure 5.4: Distribution of the relative error on the back of the hand by using the algorithm by Yoon et al. [Yoo16a] and the magnet [7.5×20 mm; N52]. A maximum relative error of 6 % can be observed at the far end of the interaction area. The visualized results were derived from the tracking experiment. Marks on the right and bottom can be used to derive the absolute position error.

5.4 Prototype

In the previous evaluation we concluded that a cylindrical magnet [7.5×20 mm; N52] is able to track the position over the back of the hand with a maximum relative position error of 6 %. The found dimensions fit into a conventional pen, which is important to achieve a natural writing experience. In addition to the small form factor, accurate interaction is also required. For this purpose, a touch sensor (AT42QT1010) on the PCB is connected to the pen tip, which communicates its state to the smartwatch via Bluetooth Low Energy (version 5) upon contact with the skin. The round shape of the pen tip and the gold plating ensure a comfortable touch experience and an optimal contact with the skin even in an inclined position. Figure 5.5 shows the individual components.

A 6-DOF IMU (BMI160) is integrated to measure the orientation of the stylus, which requires an initial calibration on first use. For this purpose the prototype must be held perpendicular to the back of the hand for one second. Based on the sensor data measured at 100 Hz, the position is then calibrated with Madgwick's AHRS [Mad11]. A magnetometer (BMM150) was also installed for test purposes, which cannot be used for orientation estimations due to the strong magnetic field. This leads to an accumulated drift over time, which, however, does not noticeably influence the accuracy due to the low-noise sensor data over the short duration of an interaction [Yoo16a]. To alleviate such tracking drifts, an indirect mapping is applied which does not require further calibration. In addition, it is thereby not necessary to further adjust the tracking area to the size of the hand.

The 7.5×20 mm magnet is mounted at a distance of 15 mm from the pen tip, also serving as a pen handle that restricts the degrees of freedom in which the pen is held.

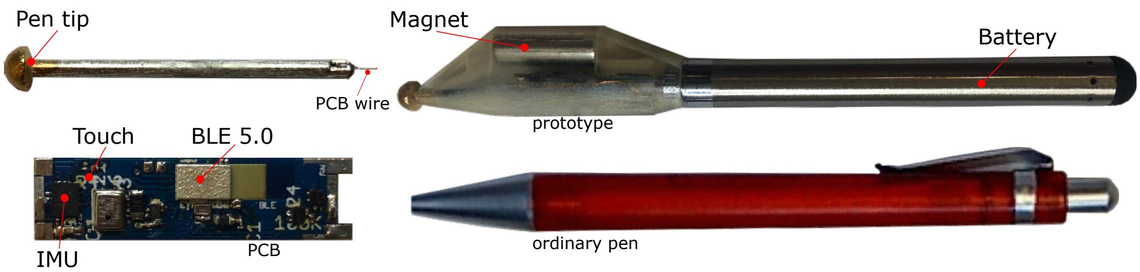

Figure 5.5: Prototype PentelliZen in detail: On the right PentelliZen is compared to an ordinary pen. With a length of 14 cm and a diameter of 9.5 mm it does not substantially differ from ordinary pens. The [7.5×20 mm; N52] magnet is integrated into the grip with a dimension of 9.5 mm×16.5 mm. The left side shows the internal PCB and the gold-plated pen tip with a diameter of 6 mm. A soldered wire on the pen tip is connected to the PCB [8 mm×30 mm] for recognizing input events via the touch sensor. The battery can be detached for charging.

This is advantageous as the corresponding yaw angle of the prototype can only be calculated using a magnetometer. In addition, the pen's center of gravity is shifted further towards the tip, which positively affects the writing comfort. In total the prototype has a weight of 18 g, of which 12 g are used for the 180 mAh battery. It can be continuously operated for 12 hours without recharging the battery.

For demonstration purposes an example painting application was implemented on the *Huawei Watch 2*. A cursor on the display shows the current position. As soon as the pen tip touches the back of the hand, a line is drawn.

5.5 User Study

A user study was conducted in order to compare the magnetically tracked pen to direct finger input in precise pointing and drawing tasks. Twelve right-handed volunteers aged 18 to 52 years (mean age $\bar{x} = 29.0\ years, \sigma = 11.7$) were recruited, including four women and eight men. One female and five male participants owned a smartwatch, which they use on a daily basis.

At the beginning each participant was introduced to the topic of the study. After filling out a demographic questionnaire, the participants had the opportunity to familiarize themselves with the drawing application and the pen-based interaction approach. The watch was comfortably but tightly fastened on the wrist. Since we use relative position tracking, a displacement of the watch could shift the tracking area. Although we expect this as a rather minor effect, we excluded related confounding variables by a tightly fastened smartwatch. The participants had to sit in front of a table and could rest their arm on the table at will. The experimenter started the application and the study ether began by using the index finger or using the pen prototype. Since we use the anthropometric measurements of the back of the hand [Aer95] and relative position control, no calibration had to be performed. We demonstrated to the participant how to move the cursor on the watch display by moving the pen across the back of the hand. When using the pen, the participants first practiced drawing strokes until they felt familiar with the interaction.

The experiment with the independent variables pen input and finger input was divided into three parts with the dependent variables error rate, time and trajectory. At first, a pointing task was conducted to evaluate the general pointing performance. For this purpose, white circular targets with a radius of r=10 px (1 mm) were displayed on a black background at random positions on the display. At the edges of the display, the test software ensured that the points were always completely visible. The participants started randomly either by using the pen or the index finger to select the randomly occurring circles with 30 trials for each each condition. When using the pen, a grey dot cursor with a radius of r=5 px (0.5 mm) was displayed. In order to quantify accuracy and speed the deviation from the center of each input event and the interaction time between two events were recorded. Figure 5.6, left, shows the experiment when using the pen. The white target circle had to be selected by moving the light grey dot cursor over the target and touching the back of the hand with the pen tip. When using the index finger, participants had to directly touch the white dots without showing a cursor. The test system recorded accuracy and time. We did not record the entire trajectory of the cursor, as it is incomparable to finger input in this experiment. The accuracy was calculated by the distance in pixels between the center of the white target circle and the position on the display where an input event was triggered. The cursor stayed visible all the time when using the pen, except it was moved out of the display dimensions to make the interaction similar to a touch pad. When starting the first experiment, the cursor was placed in the center of the screen. The time was measured between two subsequent input events. Then the previously selected target was removed and a new white point was immediately displayed at a random position on the display.

Second, a list scrolling experiment was performed to determine the applicability of the technique for menu selections and other future fields of application, such as web browsing. The participants had to select a given name from a randomly arranged list either by using the prototype or the index finger. Times and pointing errors were measured for 10 trials. Nine names were displayed on the screen at once in the vertical list. The list consisted of 20 names in total. To scroll through the list with the pen, the participants had to place the pen tip on the back of their hand and stroke it over the skin. The same interaction had to be performed with the index finger by swiping vertically over the display. During execution a scroll bar was visible at the right corner. Figure 5.6 on the right visualizes the experimental procedure. The task started with displaying the name until the participants touched the pen tip or the display with their finger. Then a randomized list of names was displayed, in which the target had to be selected. When a selection was executed, the next randomized name was displayed.

Finally, a drawing task had to be carried out. The participants traced different shapes as accurately and as quickly as possible with the index finger or with the prototype. In random order, handwritten letters (M, ß) and periodic rectangular functions with three levels of difficulty (i.e., with three, four, and five repetitions) were drawn. The templates are shown in Figure 5.8. From each drawing, measurements of time and position were recorded in order to evaluate whether the prototype offers an appropriate alternative to conventional finger input. The letters were chosen such that they represented a range of different shapes and number of strokes. We wanted the participants to draw familiar shapes of letters without lifting the pen. The *ß* is suitable because it contains round shapes while the M consists of connected straight lines in different drawing directions. The rectangle templates are easy to draw and to compare with each other. At the end of the experiment, a final questionnaire was handed out to gain insight into the individual thoughts of the participants on this pen-based input device.

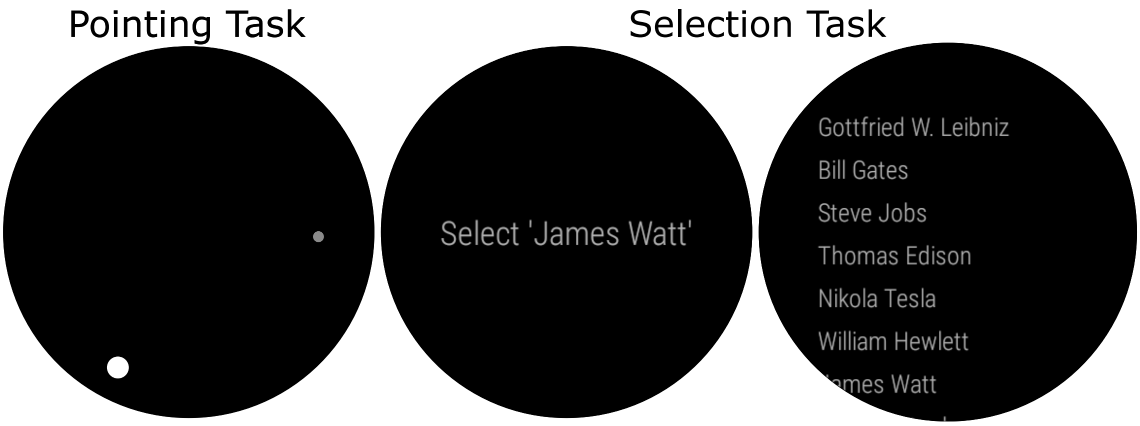

Figure 5.6: User study tasks: The participants started with a pointing experiment and continued with the selection of a random name from a vertical, randomized list of 20 names.

5.5.1 Results

All general results of the experiments are visualized in Figure 5.7. The pointing task shows that a slower but more precise interaction can be performed with the pen prototype. According to a t-test, there is a significant difference in error rate between pen (M=2.6 mm, SD=1.6 mm) and finger (M=3.1 mm, SD=1.5 mm), $t(12)=2.48$, $p<0.02$. Likewise, there is a significant difference in pointing time between pen (M=1.4 s, SD=0.56 s) and finger (M=0.61 s, SD=0.24 s), $t(12)=8.37$, $p<0.0001$. Four participants mentioned that the task was relaxing and similar to playing a game when using the prototype.

The list scrolling task shows similar results as the pointing task. In this task names had to be selected from a vertical list. Selecting another name than the requested one was counted as an error. In six cases the participants selected a wrong name from the list with the pen compared to seven errors when using direct input with the finger. According to a t-test, there is a significant difference in selection time between the finger and the pen. The finger (M=3.99 s, SD=2.5 s) achieves a significantly faster selection time than the pen (M=7.2 s, SD=3.8 s), $t(12)=5.14$, $p<0.0001$.

Last, the applicability of the pen for drawing applications was compared to direct finger input. In addition to tracing a “rectangle signal”, two letters had to be traced using templates shown on the display. A t-test was performed on each pair and the results are noted in Table 5.1. For each drawing type, there was a significant difference in the drawing time between pen and finger according to a t-test. For comparison of the drawings to the template, we calculated the corresponding DTW distances of the drawn path to the respective printed template shape. Both letter templates as well as the most difficult rectangle template with five repetitions showed a significant difference according to a t-test. For each template, DTW distances of

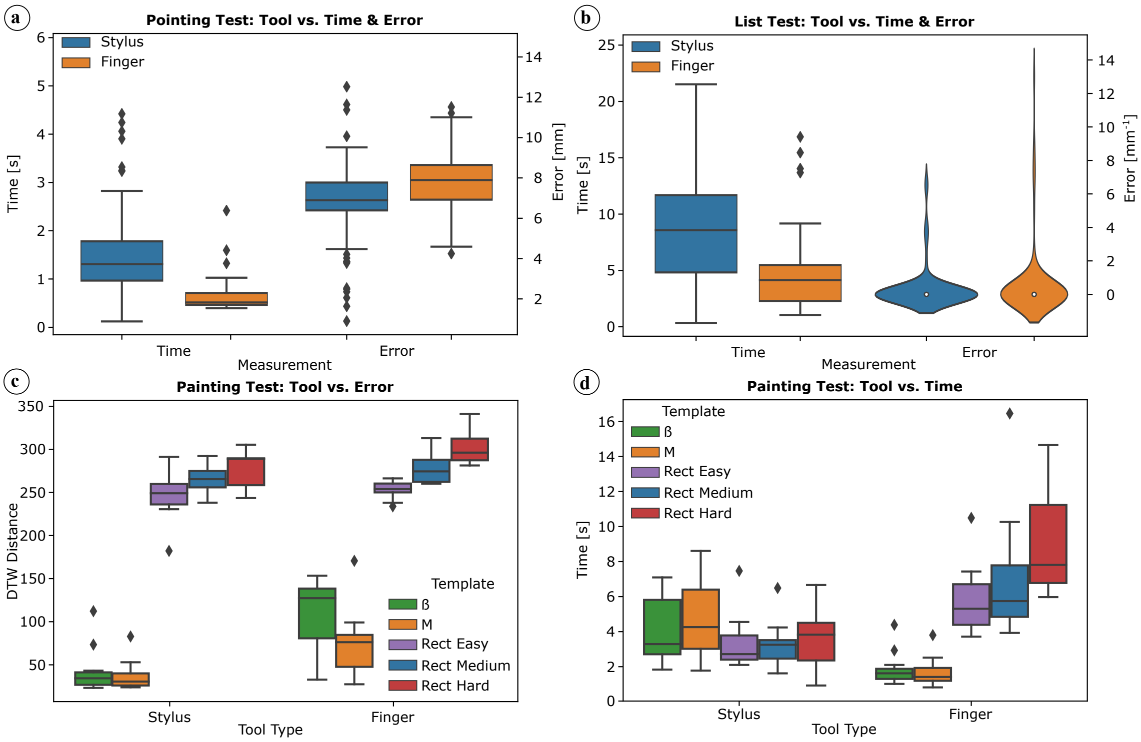

Figure 5.7: User study results: (a) shows that pointing tasks can be performed more precisely with the pen, but that it requires more time than typing with the finger, the list task shown in (b) confirms this previous observation, (c) visualizes the precision of drawing templates, and (d) the corresponding drawing time.

finger and stylus drawings were compared. The medium difficulty rectangle template with four repetitions showed a weak difference and the easy rectangle template showed no significant difference. All values are shown in Table 5.1. To exemplify, Figure 5.8 visualizes the results of one participant.

The results show that using the pen leads to more precise drawings than the finger, but requires more time. Since none of our participants had ever used a similar device on the back of the hand before, it is to be expected that, after a longer period of use, the task completion times will decrease.

A final questionnaire was provided to get insights into the participants' subjective reactions to the proposed technique. The results are shown in Figure 5.9. In general, a comfortable writing experience is achieved, with three participants remarking that they would not be willing to carry around a pen with them. The interaction was perceived to be precise, but four participants mentioned that touch events were less precise with rough or dry hands. The applied indirect mapping of the cursor position to the stylus was perceived to be a natural design similar to a touch pad. Mostly the participants perceived the interaction to be slower when using the pen.

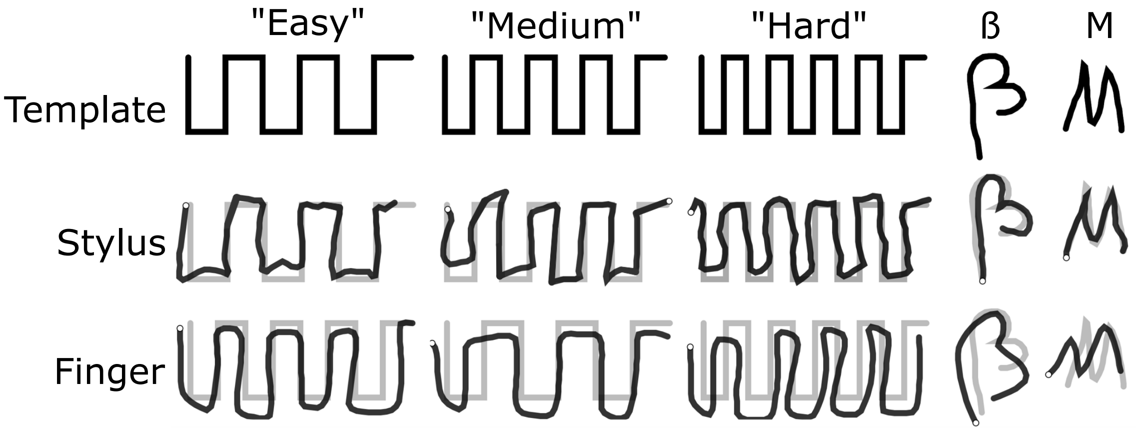

Figure 5.8: Templates of the drawing task: The drawings are taken from one participant of our study. It is apparent that the drawings created with the stylus (second row) are much more precise than the finger-based drawings (third row). Particularly in finer structures, differences become evident.

Table 5.1: Results of the t-tests of the drawing experiment: In general, the pen is more precise than the finger when drawing fine structures. Drawing and writing with the finger is faster in all cases. The results are based on twelve participants.

Template	Method	Precision in [DTW dist.]			Time in [s]		
		Mean	SD	p-Value	Mean	SD	p-Value
ß	Stylus	42.2	25.9	0.0001	4.64	2.2	0.0003
	Finger	109.5	38.64		1.68	0.85	
M	Stylus	74.9	38.5	0.0049	4.1	1.83	0.0009
	Finger	240.6	31.6		1.85	0.95	
"Easy" rectangle	Stylus	253.2	10.1	0.3523	5.7	1.9	0.0029
	Finger	245.7	25.8		3.4	1.5	
"Medium" rectangle	Stylus	265.2	15.8	0.0942	7	3.5	0.002
	Finger	277	16.9		3.3	1.3	
"Hard" rectangle	Stylus	278.9	20.6	0.0099	8.9	2.8	0.0001
	Finger	301.4	18.2		3.6	1.6	

Two participants stated that they would expect a faster interaction in the long-term and that they could imagine increased productivity when using the pen on their smartwatch. We also asked our participants about potential future fields of application. They mentioned signature identification and handwriting recognition most often. But they also saw drawing as a suitable application for such a device. Similar tasks appear in productivity applications, i.e., email, messaging, or notes, resulting in more engaging interactions in these scenarios. Some participants were concerned about the small display for tasks like web browsing and assumed games as a more suitable use case. All participants found the light-gray cursor with a radius of $r = 5\ px$ (0.5 mm) to be appropriate and clearly visible.

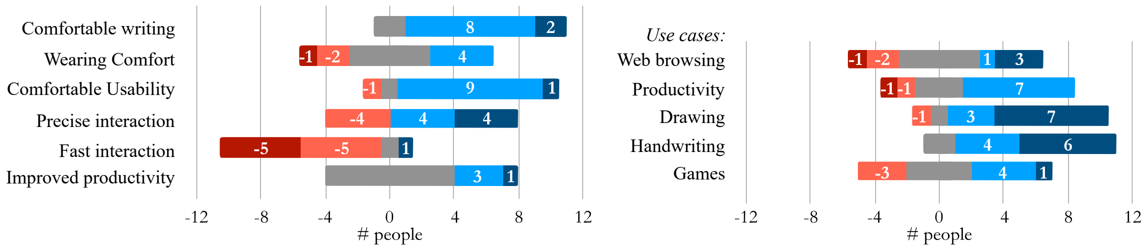

Figure 5.9: User study questionnaire: The participants found the prototype to be comfortable and easy to use for drawing and handwriting applications. The small display size was seen critical for tasks like web browsing.

5.6 Drawing Application

To maximize the drawing area, settings are accessed by moving the cursor to the edges of the round smartwatch display in each of the four cardinal directions. The menu items are invisible until the cursor is moved near the edges of the canvas view. A small arrow icon occurs on the top of the screen and serves as a cue to indicate presence of the menu. The menu is invoked when the cursor is near the icon and the user touches the pen tip. To avoid unintended input, this functionality is hidden and disabled when drawing a stroke into the vicinity of the edges.

In the canvas view, scroll bars have been implemented at the bottom and right edge, which adjust the viewing area in x and y direction by moving the pen and touching the pen tip. On the left display corner, the field of view can be adjusted in z direction. Changes are applied as long as an input event is detected.

The menu is accessed at the top of the canvas view. When accessed, the menu can be exited by moving the cursor to the top again and touching the pen tip. Menu tabs are accessible when the cursor is moved towards the bottom display edge. Since we only use a limited number of settings, only two menu tabs were implemented. Due to the small display size we expect that users primarily use the app for simple and spontaneous sketching. Hence, the full functionality of an desktop PC drawing application would overwhelm the interface and lead to unnecessary complexity.

The view can be locked by a U-lock icon in the menu. The symbol also indicates the current status of the canvas. Arrows indicate undo and redo functionality. When touched, the canvas view is updated according to the action. An eyedropper tool has been implemented to pick colors in the image. When selected, the canvas view is opened up with a color picker cursor indicating the current action. By opening the main menu at the top of the screen again, the task can be canceled. Stroke width and transparency icons open up a vertical slider view that updates the current stroke setup. By moving the pen vertically up or down, adjustments can be made. Changes are processed by touching the pen tip. The dashed lines icon is used to delete strokes in the canvas. When selected, the canvas view is reopened with the background color for drawing.

Three-dimensional interactions in the air require to touch the pen tip with a finger as visualized in Figure 5.10. Preliminary tests have shown that this action can be easily performed, potentially requiring a slight adjustment of how the stylus is held. The pen position in the air is always mapped to a 2D position on the display, provided that the device is in the sensing range of the smartwatch. Three in-air interactions have been implemented.

When selecting the magnifying glass, areas of interest (AOI) can be enlarged according to the position over the hand. By touching the pen tip with the cursor over the magnifying glass icon, the canvas view is opened with a rectangle cursor that visualizes the zooming area. By lifting the pen, the rectangle is enlarged. When touching the pen tip, this selected area of the canvas is enlarged to the display size. The normal cursor is displayed again and drawing can be continued. Various icons, sorted by categories – which are taken from the material design scheme [Goo20] – can be added to the drawing by placing them on the canvas. They can be enlarged using the described in-air lifting technique. When an icon has been selected from the library, the canvas view is opened and the icon is shown instead of the cursor. Its size can be adjusted by lifting the pen. The action can be aborted by moving the cursor to the upper display edge. Selecting the color palette icon in the menu opens up a color picker view. Via two-dimensional cursor movement, the color (hue and saturation) are specified. When lifting the pen, brightness (value) can be adjusted. Colors are selected by touching the pen tip.

5.6.1 Usability Study

A qualitative usability study was conducted to evaluate the suggested pen-based interaction techniques on the back of the hand (on-skin) and above the hand (in-air). Five volunteers were recruited (mean age $\bar{x} = 21.8\ years, \sigma = 0.84$) including one male smartwatch user.

After an initial questionnaire with demographic questions the participants put on the smartwatch comfortably but tightly and explored the preliminary drawing application from the first study. They were introduced to think aloud while trying

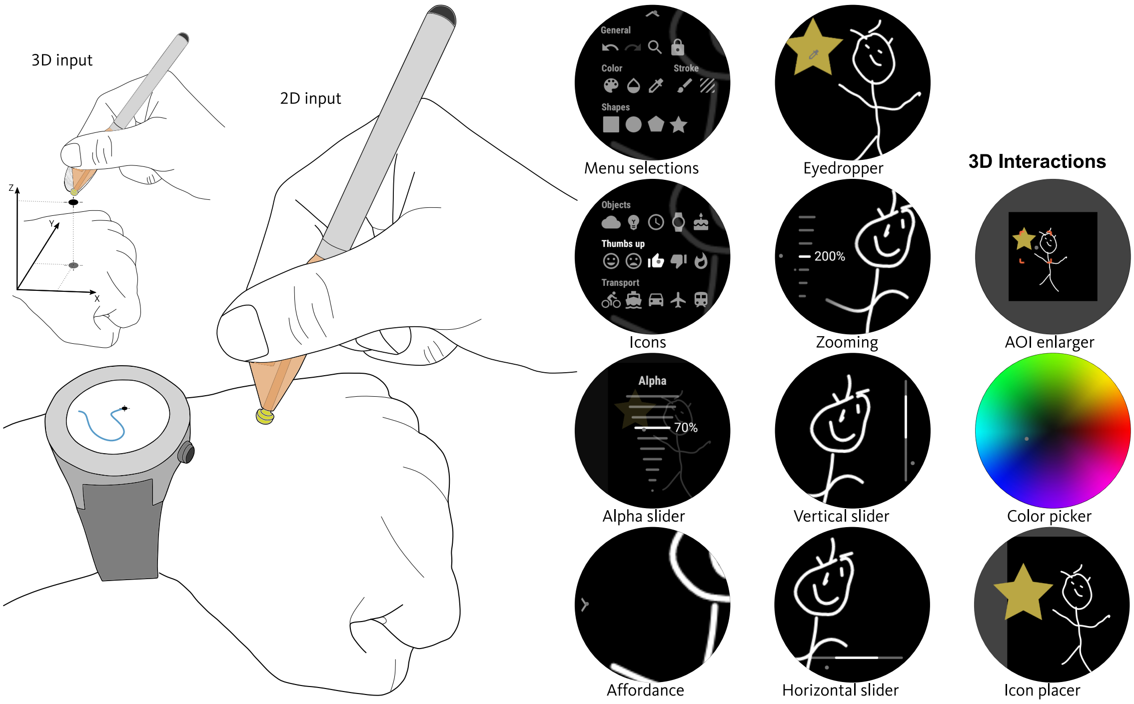

Figure 5.10: Drawing application: on the left 2D (on-skin) and 3D (in-air) interactions are visualized. By touching the pen tip with the finger in the air (upper left), selection events can be issued. On the right, different screen shots of the application are presented.

the application. Once they gained sufficient experience to successfully draw lines on the canvas, the extended application was started. Again, the participants were told to think aloud while exploring the functionality. The experimenter noted down their thoughts without interventions. When the participants had tested all of the settings, they could take off the watch. In case an option had been omitted, the experimenter assigned additional tasks until all of the features had been explored.

Subsequently, a questionnaire was handed out, starting with general questions about the hardware prototype. Then the participants answered usability-related questions regarding the final application. We wanted to know whether they were able to interpret menu icons and the current state of the application. The next series of questions was intended to assess whether the presented pen prototype is a good addition for everyday use. At the end, specific questions were asked regarding the implemented interaction techniques and software.

5.6.2 Results

The results of the usability study are shown in Figure 5.11. Since only one of five participants stated to wear a smartwatch in everyday life, there might be a negative bias in the subjective answers. However, all participants used mobile devices in everyday life. Therefore we assume that all were able to get used to new technologies.

Generally, the participants felt comfortable about interactions on the back of the hand, but nevertheless could not imagine using this device in everyday life. Two participants found drawing on the back of the hand to be useful and all participants agreed that the virtual cursor together with relative position control mapping is appropriate for drawing on smartwatches. However, the results show a rather negative pattern. The single smartwatch user stated that without a well-integrated commercialized product, he would not use such a pen with his smartwatch. He remarked that the entire watch user interface – i.e., all of its applications – should be designed for use with the pen. Three participants found the display to be too small for sketching.

The second part of the questionnaire focused on the application, which confirmed that the app had been designed in a suitable way for pen-based operation. Due to focusing on the primary functionality, the complexity had been intentionally kept low. Since the target application area that we originally had in mind was rather casual drawings and sketches, as for messaging a personalized message to a friend, a full-blown drawing application would not have been suitable. One participant remarked missing features like cut, copy, and paste.

The interaction techniques were perceived to be intuitive. The participants stated that the 3D input and the scrolling technique should be improved. Scrolling was less intuitive for the participants because the role of the edges of the display was not clear initially. A better option might be to show the related scroll bars when the cursor is

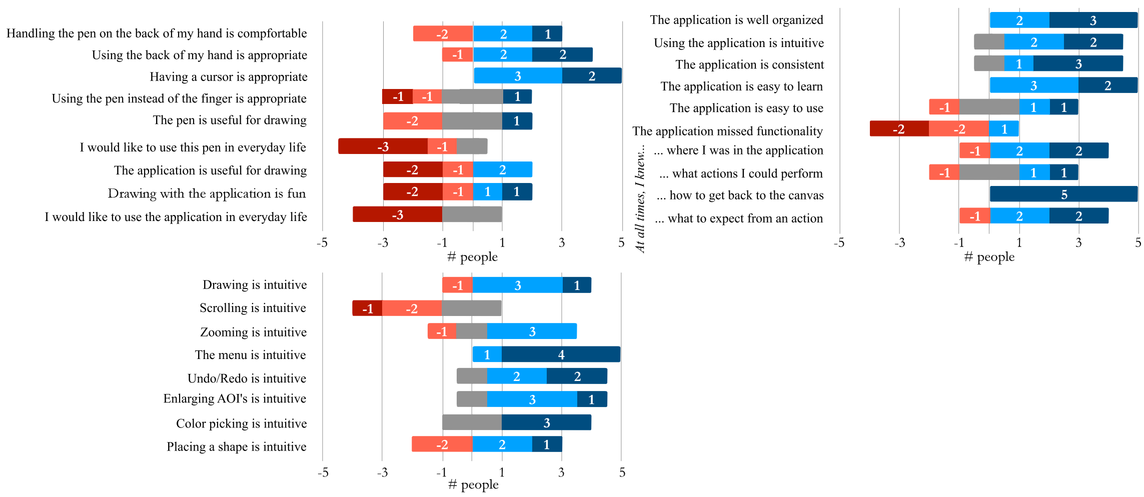

Figure 5.11: Usability study questionnaire: The participants judged the 3D interactions as less intuitive for placing canvas elements, but as well suited for adjusting parameters like colors.

moved to the edges. In analogy to a touch pad, scrolling could also be implemented at an absolute position on the back of the hand at maximum distance from the drawing space. This would require an additional calibration to the dimensions of the user's hand. While exploring the application, the experimenter noted challenges when 3D interactions were performed. The participants had to apply shape placements twice as they did not intuitively understand how to lift the pen for resizing. Mostly they wanted to touch the back of the hand to interact with the canvas and were confused why the icon was getting smaller when the pen was moved towards the back of the hand. One participant stated that mixed 2D and 3D interactions lead to a less intuitive drawing application. Three participants said *"Uh-huh"* after completing a 3D task, which indicates that 3D interaction are challenging and should first be tried out before they are fully understood.

5.7 Discussion

Our experiments revealed that 3D tracking in the dimensions of the back of the hand achieve a maximum relative tracking error of 6 %. The preselection of a suitable magnet was necessary, otherwise errors in the position determination may occur. We used Newton's method for solving the equations in order to minimize the computational effort. Using the bisection method could might improve the tracking performance, but is challenging to apply on current smartwatches. We have noticed an increase in battery consumption and temperature on the watch when using the prototype. Although the code has been optimized, more efficient processors would thus be advantageous. An enhanced sensor sensitivity could increase the tracking accuracy and allow to use smaller magnets. Our proposed system was able to achieve more precise but slower interactions compared to direct finger input on the tiny display. It has been shown that direct input on the back of the hand is understood more intuitively than in the three-dimensional space around the back of the hand. These results can be indirectly transferred to other applications for smartwatches. The use case of a drawing application applied here was chosen as an example scenario. For example, it would also be possible to implement a web browsing app or simple games such as whack-a-mole and hot wire.

The analogy to a touchpad on the back of the hand is easily understandable. However, the dimensions of the tiny display limit the capabilities. Thus, we assume the presented technology could support applications on other mobile devices such as tablets. For instance, interaction elements like color pickers or zooming from drawing applications could be outsourced to the smartwatch. It should be noted that the conductivity of the skin is user-dependent and therefore a calibration would be advantageous. We experienced challenges with rough or dry hands and the sensitivity of our touch sensor. While lines were drawn on the back of the hand, we observed individual samples in the sensor stream that did not recognize input events. These samples can easily be discarded by filtering. Further increasing the capacitive touch sensitivity can lead to falsely triggered input events when the pen is held in hand. We have therefore used a shielding around the internal parts of our touch circuit which can be further improved in future prototypes.

5.8 Conclusion

We proposed an interaction technique and device for drawing on the back of the hand for ultra small wrist-worn displays. By integrating a magnet in a stylus, the position relative to a magnetometer can be determined with high precision. Different dimensions and magnetic field strengths were tested with the result that a magnet with a dimension of 7.5×20 mm and a magnetization degree of N52 achieves acceptable results on the Huawei Watch 2. Furthermore, different algorithms were evaluated regarding their accuracy in the interaction space, which showed that the 3D position determination is significantly more accurate. An initial user study accompanied by a usability experiment served to evaluate the field of application. It showed that by using the stylus, more precise interactions can be made than with the finger. Contrarily, the interaction time increases with the pen. However, the study shows that for applications like drawing on a tiny screen, direct finger input leads to higher error rates and higher task completion times than indirect pen input on the back of the hand. The analogy to a touchpad on the back of the hand was easily understood while 3D interactions required additional learning effort. In the future the presented technology could be used to outsource interaction elements from drawing applications on a tablet to the watch for maximizing the drawing space on the tablet.

CHAPTER 6

Tracking Surfaces in the Wild by using Audio and Motion Data

The previous chapters focused on interactions with pens. We identified surface scratching sounds as supportive features to the motion of the pen which contain general information of the task as well individual features that can identify writers. We could also substitute camera systems for sketching by using magnet tracking to get a relative position to wrist-worn devices. The skin on the back of the hand was introduced as an ubiquitous surfaces that is capable of being used similar to a touchpad. This chapter evaluates whether scratching sounds combined with motion data can identify surfaces in the wild. For this purpose, we analyze two use cases where surfaces play an important role. At first, we introduce audios as supportive features to track cycled road conditions and then continue by analyzing snow types on ski slopes.

Do scratching sounds contribute to recognize surfaces in the wild?

Do recognized surfaces support users in their route selection?

This chapter is based on the master thesis by Timon Bressgott [Bre20] and a poster at *MUM 2020* [Sch20b]. The topic of Timon's master thesis was developed in cooperation with the *Institute for Cartography and Geoinformatics* (IKG) at the Leibniz University Hannover for the *Mobiler Mensch* (MoMe) project. Accompanying the topic, we held a workshop at *MobileHCI 2020* [Sha20]. Furthermore, the used hardware from this project was tested for measuring skiing slope conditions at the *Winter School 2020*. In collaboration with Jonathan Liebers and Stefan Schneegass from the *University Duisburg-Essen* tests were conducted and subsequently published as a poster at *MUM 2020*.

- Schrapel, Maximilian & Liebers, Jonathan & Rohs, Michael & Schneegass, Stefan: 'Skiables: Towards a Wearable System Mounted on a Ski Boot for Measuring Slope Conditions'. *19th International Conference on Mobile and Ubiquitous Multimedia*. MUM 2020. Essen, Germany: Association for Computing Machinery, 2020: pp. 320–322

- Shadan Sadeghian & Maximilian Schrapel et al.: *Workshop on Designing Technologies for Future Forms of Sustainable Mobility. MobileHCI'20, Oldenburg, Germany.* 2020

6.1 Introduction

Wearable technologies and the Internet of Things (IoT) are becoming increasingly popular. They weave into a variety of application areas which can be divided into health, activity recognition, localization, safety and other special purposes such as education or law enforcement [Dia20]. Wearables can collect and share data to monitor a user's state and support them in smart decision making. A seamless communication among these inconspicuous small devices enables to make a user's environment smart. Their ubiquitous availability partially realizes Weiser's vision of future computing [Wei91].

Recent years have shown the great potential of wearables in physical activities starting from beginners to professional athletes for improving their skills [Cam18]. Outdoor activities often involve unpredictable risks related to the dynamic fashion of an environment. For instance, a large proportion of bicycle crashes result from unpredictable road conditions [Hee11]. Measuring and providing current road conditions is an enduring research topic. Until today, public available maps contain no complete data on the road conditions. Since sensor-equipped bicycles have found their way into the cities, ideas have emerged to use their data towards the vision of smart city monitoring [Du19]. Especially due to the emergence of new forms of micro-mobility on the last mile, road conditions are becoming increasingly important. These novel vehicles are often equipped with a variety of sensors that can measure different properties of an environment. In addition to micro-weather measurements [Cas14], air pollution [Fie21; Sav10] and ambient noise monitoring [Lax19], road quality measurement is an important aspect [Ant93] that relates to safe operation [Wag20] and cycling speed [Bou18; Höl12]. Road surface conditions are dynamic and weather dependent [Eri08]. An asphalt road can be well suited for cycling in summer but involve unpredictable risks in winter due to black ice [Win10]. This exemplifies that the smoothness of a road alone is not sufficient to determine the suitability of a path for cycling. There are even greater challenges in winter sport activities. Winter resorts categorize their slopes according to levels of difficulty. However, slope conditions are highly dependent on the weather conditions and the daily traffic. A slope categorized for beginners can involve unpredictable risks due to snow bumps or icy passages. For instance, most injuries occur when snowboarding on ice and packed snow [Bia95].

Cycling and winter sports like skiing have in common that road or slope conditions have an influence on the risk of injuries. In both cases, it is challenging to create a smart environment that supports the user in route planning processes or warns the user of dangerous passages. Sensor networks placed on the route would require high maintenance and are associated with expenses. Therefore, approaches where sensors measure the surfaces directly on the bike or on the skis are more suitable. Both the tires and the skis rub on the surface during the activity, what produces noises that can provide information about the condition of a surface type. Likewise, the vibrations on the surface contain information that could help identify a condition.

In this chapter, we analyze whether a fusion of audio and motion data measured on sport equipment can recognize different surface conditions. For this purpose, we introduce a sensor unit that can be attached to the bike at the seat stay or at a ski boot. Measurements are sent via Bluetooth to the smartphone for further analysis. In a study with 17 cyclists we collected a dataset of 130 km of cycling on different bikes. We compare our approach based on convolutional neural networks to commonly applied road roughness clustering [Wag20] and show that it is crucial to consider tire pressure. By surveying our participants, we show that bike route preferences are weather dependent. Further, we conducted an experiment with an experienced skier in an winter resort to test our approach for identifying snow conditions on the slopes. In a subsequent online survey of skiers with different skill levels we show that skiers prefer several snow conditions on the slopes.

6.2 Related Work

In this section, we briefly present a selection of works that address safety and road conditions for cycling and winter sports.

6.2.1 Cycling

First research in the 1980s analyzed route choices of cyclists where facility type, traffic level and surface types were identified as main criteria for pleasant rides [Bov85]. During that time the International Roughness Index (IRI) was established to offer a standard road condition measure for vehicles [Say86]. Over the years, further road roughness measures [Ale17; Bíl15] and anomaly detectors [Eri08; Wij18] were proposed. They have in common to use accelerometer and GPS measurements to categorize road profiles according to their roughness. Bil et al. established the Dynamic Comfort Index (DCI) to rank cycle paths according to their ride comfort [Bíl15]. However, a good DCI value does not necessarily indicate an ideal route choice for cyclists [Ehr12]. Even navigation systems do not fully consider all facts in their route planning algorithms [Gol95; Joh17]. Off-street paths play an important role in the cycle route choice, as cyclists tend to avoid high traffic volumes [Bro12; Win10]. More generally, cyclists are very sensitive to aspects of safety and comfort [McC15; Win10], where the quality of road surfaces is more important than other measures, such as hilliness or delays caused by traffic stops [Sti03]. Especially with electrical power support, distances and slopes are of less importance [Zha21a]. Cyclists are willing to travel up to additional 20 minutes for having pleasant and safe rides [Til07]. Poor road qualities correlate with an increased risk of fault events [Fei20].

Bike lanes, like all other roads, are subject to constant change, e.g. due to different weather conditions and abrasion. This dynamic fashion requires continuous measurements and motivates some cyclists to monitor and share data to publicly available maps like *OpenCycleMap* [Ope20]. Mixed conditions and especially snow are a major challenge for many forms of sustainable mobility [Ber03; Giæ98]. Road maintenance is a key factor, not only during winter season, to encourage cycling instead of driving [C4020]. Smartphones mounted on the handlebars can measure and share standard roughness measurements [Eri08; Kar20; Wag20; Wij18; Zan18; Zha16b] and recognize road signs with the rear camera [Rei17]. Their ubiquitous availability [Bal06] offers the possibility to reach a large number of users. Permanently mounted systems [Cas14; Fie21; Höl12; Lax19; mei20; Mia20; Sav10; Tay10; Wu20] additionally offer the opportunity to continuously measure weather [Cas14] or air quality data [Fie21; Lax19] while parking the bike.

Audio measurements have been used for ambient noise monitoring [Lax19; Sav10; Tay10] during cycling. To the best of our knowledge, there has been a lack of research on combining vibration measurements with audio signals to estimate road conditions on bicycles. Audio measurements have also been used to identify wet roads for aquaplaning warning systems in cars [Alo14], on robots [Zür19], and, in combination with motion data, different surface conditions have been recognized while walking [Mit19]. It has not yet been analyzed whether a sensor fusion approach can increase the recognition rate of road surface types and conditions of cycled paths.

6.2.2 Winter Sports

Winter sports equipment offers increasingly higher safety through the latest innovations [P H00]. Further, the more frequent use of helmets resulted in a lower proportion of head injuries [Ack07]. However, wearing a ski helmet can significantly reduce the ability of spatial hearing. Niforatos et al. attached LIDAR sensors to the back of the ski helmet to signal, via LEDs at the front, when skiers approach from behind [Nif16; Nif17]. Displaying personal information on ski maps via head mounted devices can also increase safety, as hazards can be marked on the routes [Fed16]. Combining augmented reality with extending peripheral perception was proven to be a promising field for future applications [Nif18].

Wozniak et al. identified general guidelines to support outdoor activities [Woz17]. According to their findings, obstructing the outdoor experience should be avoided and route planning, including risk management, should be supported. Displaying piste maps including weather and slope conditions on a mobile device requires continuous updates [Dun07]. Entering information manually into a database can impair the outdoor experience and produce erroneous data. Further, Wozniak et al. also identified that skiers' groups prefer to exchange entered information only among their group members [Woz17]. According to Fedosov and Langheinrich location, media, and reference data are the most-shared information between groups of skiers [Fed15].

The skiing experience level within a group can greatly differ [Woz17]. Experienced skiers mostly support beginners, but also feedback systems can help to automatically enhance the performance of beginners on the slopes.

Body position control is of utmost importance during skiing. Hasegawa et al. added force sensors to the skis and transformed the data into acoustic signals to provide posture feedback [Has12]. However, real time audio feedback on lateral displacements and speed can be distracting and affect safety [Kir09]. Subsequent feedback can support professional skiers to increase their skills. Brodie et al. used fusion motion capturing with 13 inertial measurement units (IMU) consisting of accelerometers, gyroscopes and magnetometers attached to the body and skis for measuring posture information on a slalom course [Bro08]. Tactile feedback can provide instant assistance without sound distractions. This is very important, especially for visually impaired skiers [Hal17]. Spelmezan attached two vibration actuators laterally on each shoulder to correct body rotations and weight distributions of snowboard beginners [Spe12]. The design of these tactile stimuli must be able to represent body movements intuitively [Spe09]. To measure body posture during physical activities, pressure sensors in the boot sole and accelerometers on the snowboard [Hol09] or integrated into textiles, e.g. socks [Hol10a], can be applied. Alternatively, torsion can also be predicted via electromyography (EMG) [Lou84]. Besides posture correction, IMUs on the snowboard can classify tricks [Gro16] and simple turns [Hol10b]. On skis, carved turns were measured via attaching bending sensors [Yon10] or EMG [Mül94].

For successful tricks and turns optimal slope conditions are essential. However, conditions may rapidly change through the day due to changed weather conditions or high traffic volumes. As a result, skiers, in particular beginners, tend to overestimate their skills [ONe99]. For example, most injuries occur when snowboarding on ice and packed snow [Bia95]. Therefore, static maps are only of limited help for an optimal route selection. Wireless sensor networks can continuously measure changing conditions. To name a few approaches, snow depth can be measured by infrared light [Hen04] or ultrasonic sensors [Mal19]. Melting water is correlated with soil moisture, temperature and humidity [Ker12; Mal19]. Snow drift and velocity have been sensed acoustically [Chr99]. Seismic sensors [van16] have been utilized to improve the reliability of avalanche warning systems. However, these systems only provide indirect slope information, as they cannot be placed directly on the slope. Koptyug and Kuzmin measured gliding resistance on the slopes via pressure, accelerometer and audio measurements underneath the ski runs [Kop11]. Similarly, Leppävuori et al. measured vertical, cross horizontal, and longitudinal horizontal force components with strain gauge bridges buried in snow [AP93]. However, equipping a ski resort with a high number of sensor nodes can affect the beauty of the scenery [Ker12].

In contrast to related work, we explore whether slope conditions can be directly measured on the skis without instrumenting the slope. Skiers in a winter resorts could thus automatically contribute to improving safety.

6.3 Prototype

We aimed to design an inconspicuous sensor kit that can be attached to different sports equipment without interfering the experience. Figure 6.1 shows the PCB and the sensor kit attached to a bike and a ski boot. Within a PCB size of 30×8 mm we integrated a 6-DoF motion sensor (BMI 160) [mark *C*], an analogous contact microphone (MM34303-1) [mark *B*], and a wireless Bluetooth 5.0 connection (NRF-52832) [mark *A*]. We achieve a maximum transmission rate of 15 kHz for audio and 800 Hz for motion data. Since in real world scenarios WiFi networks or other Bluetooth connections influence the resulting transmission rate, we tested our prototype in noisy environments. We measured a stable communication with a transmission rate of at least 7.5 kHz for audio and 650 Hz for motion in a noisy university environment with 13 nearby WiFi networks and a line-of-sight distance of 5 m between a *Xiaomi 9T pro* smartphone and the prototype attached to a city bike. To further ensure stable measurements the battery voltage of 3.7 V is regulated down to constant 3.0 V. The batteries in mark *D* and mark *G* have capacities over 100 mAh. In operation, the prototype requires about 4 mA, which allows continuous measurements over a whole day. The microphone in mark *F* is covered with 3mm neoprene to prevent wind noise from interfering the audio measurements. When mounted, the microphone is pressed against the bike frame or the ski boot. Thus it operates as a contact microphone. To prevent short circuits from water, the entire PCB is coated with rubber.

For our tests on bikes we decided to mount the sensor kit at the seat stay near the dropouts. Since most of the cyclist's weight lies on the rear tire, this position ensures motion measurements resulting from vibrations on the surface. In addition, noise generated by friction on the surface is expected to be louder when measured close to the tires. This location also ensures that damping springs of mountain bikes do not cause attenuation of vibrations, which they do at the handlebar. A firm mounting is ensured by four screws, which also increases the difficulty of theft. When mounted, the microphone is pressed against the bike frame and the neoprene cover protects the sensor from wind and water. Although the battery can be replaced for test purposes, real application scenarios require a continuous operation without manually recharging the batteries. For this purpose, the battery could be charged either via solar or the dynamo. However, we wanted to exclude problems with empty batteries during our tests and only used fully charged batteries.

For testing the prototype on skis we decided to mount the prototype on the ski boot since attaching sensors directly on skis involves several challenges. First of all, high mechanical forces are to be expected directly on the skis [Yon10], which can lead to saturation of the sensors and severely affect the lifetime of the sensors and the prototype. Melting snow could possibly shorten circuits. Cold and fluctuating temperatures influences measurements. Similarly, an integration into the ski binding has been ruled out as snow can accumulate and solidify due to the weight of the skier. In addition, to mount the sensor unit irreversible modifications of the equipment are likely to be necessary [Yon10]. Previous work proposed adding sensors to the sole of the ski boot [Hol09]. This placement avoids effects of snow or coldness on the sensor technology. However, sweat can accumulate in the foot-bed and cause short circuits. Our chosen position makes the sensor unit concealed by the skier's trousers as well as discreet and protected from external influences. The holes in the housing are used to fixate the prototype with elastic cord.

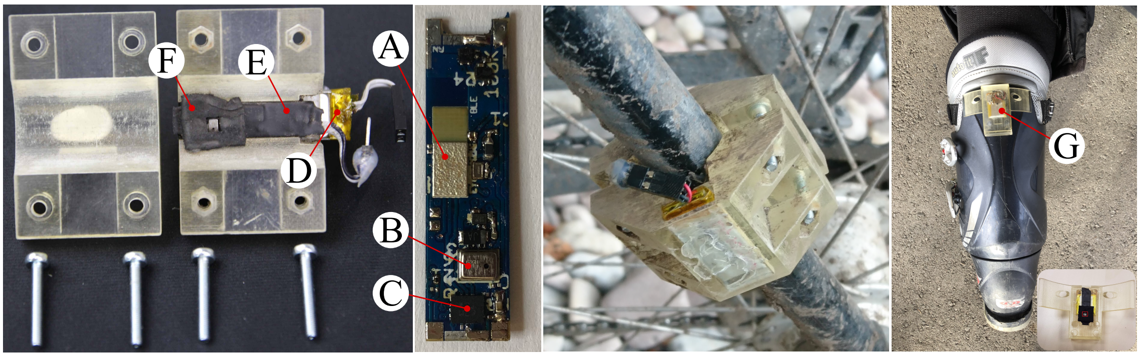

Figure 6.1: Prototype for cycling and skiing: The PCB in the middle is coated with rubber to prevent short circuits. The microphone is pressed against the ski boot and bike frame to work as a contact microphone.

6.4 User Study

To investigate whether combined audio and motion measurements can contribute to surface recognition, we conducted a field study on cycling. In order to compare our approach with related work, we also collected accelerometer data from a smartphone to compare our approach with commonly used road roughness measurements [Wag20]. In addition, one experienced skier collected a first dataset of a half day skiing. The experiment was accompanied by an online survey to collect opinions from skiers.

6.4.1 Cycling Study

The study was conducted throughout the city center of Hannover, Germany on weekdays between 11 am and 4 pm during the winter season of February 2020. Temperatures around zero degrees Celsius prevailed and no snow or rain was present during the studies. We invited 17 volunteers aged between 21 and 58 years (M = 31.1, SD = 12.0) who cycled on average 1897 km (SD = 1493 km) per year and use their bike for 4.2 days (SD = 1.8) per week. The frequency of bike types was: 13 city bikes, 2 mountain bikes, 1 racing bike, and 1 dutch bike. Our participants stated to cycle with an average velocity of $\bar{x} = 18\,{}^{\mathrm{km}}\!/\!_{\mathrm{h}}$ ($\sigma = 6\,{}^{\mathrm{km}}\!/\!_{\mathrm{h}}$). During their daily tours, they stated to mostly cycle on asphalt followed by pavement, gravel, cobblestone, and unpaved roads.

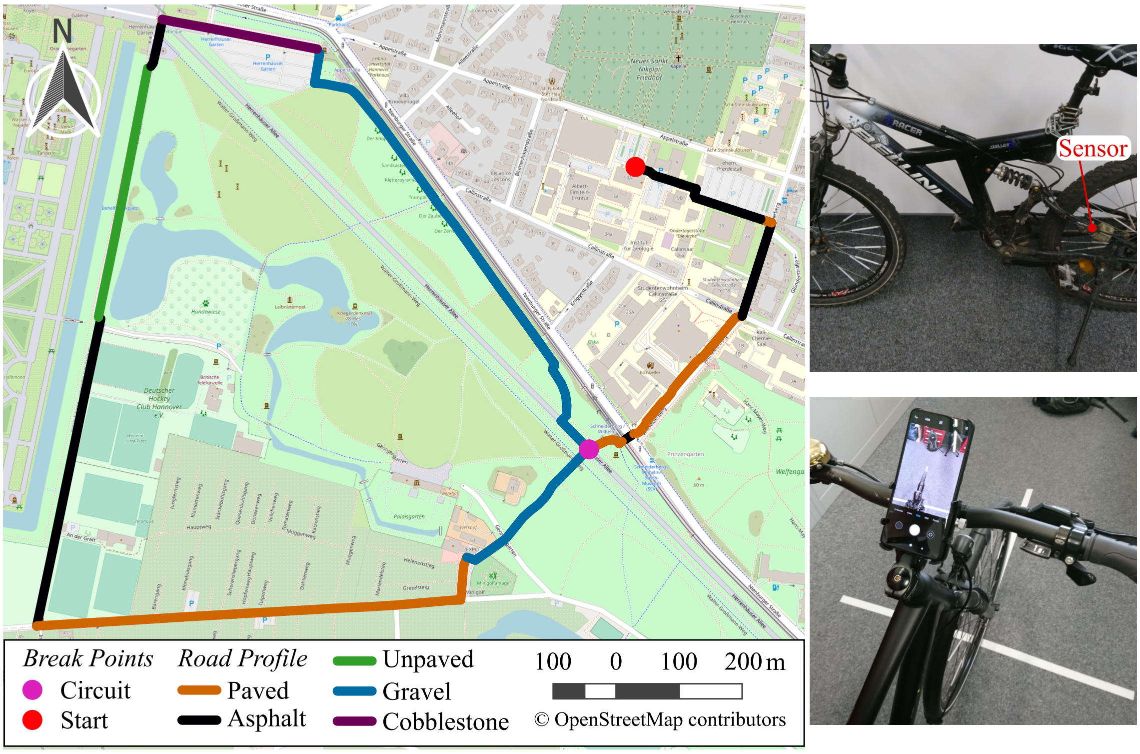

Figure 6.2: Left: Visualized route and road profiles. Top right: The mounted sensor kit on a mountain bike. Bottom right: The smartphone position. The smartphone's rear camera video was used for manual post-hoc labeling of the road type.

While the participants filled out the initial questionnaire, the experimenter attached the smartphone to the handlebars of the participants' bicycle and the *Cyclables* sensor kit to the seat stay near the dropouts. After the setup was complete the participants and experimenter put on safety equipment (helmet and west) and the study began. The cycled route, shown in Figure 6.2 left, was selected based on traffic safety and on containing a variety of common surfaces in city centers [mei20]. For post-hoc labeling purposes, we recorded videos to exclude passages from training data where the participants cycled by the side the cycle path or over a curbside. The route includes sections near a tramway line and frequently used cycle paths next to main roads as well as trails in a park. The participants had to follow the experimenter who targeted a moderate speed of 15 km/h. The circle path was driven once in each direction. The initial direction was chosen randomly. The path included a short break in between, where the experimenter checked and noted down the tire pressure of the participant's bike. The tire pressure was subsequently inflated to the optimal level for each bicycle as indicated on the tire. After heading back to the starting point, the experimenter dismounted the hardware, while the participants filled out a closing questionnaire. The entire study lasted about 60 minutes and included 7.6 km of cycling with 2.98 km untagged road profiles in *OpenStreetMap*.

6.4.2 Skiing Experiment

We decided to measure data throughout a half skiing day of a single skier. This allows us to provide first statements on the feasibility of our approach and to identify limitations for future research. The study took place at the skiing resort 'Hoher Ifen' in Kleinwalsertal, Austria on February 18, 2020. Mixed, but mainly sunny weather conditions with temperatures around 0°C (32°F) and a high number of visitors gave us a high diversity of slope conditions. One of the experimenters (male, 28 years) with multiple years of skiing experience attached the prototype at his left ski boot. For collecting data, a Huawei P30 smartphone was connected to the prototype via Bluetooth and the smartphone was then placed in the experimenter's backpack. Besides audio, gyroscope and accelerometer data, the GPS positions and timestamps were recorded on the phone as well, as on a GoPro action camera mounted on the chest. The camera was aligned so that the snow between and at the ski was recorded. The experimenter rode at normal pace on slopes of all difficulty levels. A second experimenter followed the skier all the time and both discussed and noted down the slope conditions after each run. The slopes were categorized in wet, ice, groomed, packed, flat, mogul, and powder snow. The classes were chosen based on Müller's categorization to distinguish biomechanical characteristics in swinging techniques during alpine skiing [Mül94]. In total, over four hours of skiing data were recorded, starting from 8:30 pm to 12:45 pm, including transportation and waiting for the ski lift. All slopes of the resort have been covered in this process.

The aim of the subsequent online survey was twofold. First, we asked the skiers towards their opinion on our idea and if they would benefit from more detailed and more dynamic map information. Second, we also wanted to investigate whether recorded videos during the feasibility study can be used to support or automatize labeling, so as to extend the data set in the future. We recruited 17 volunteers (10 male, 7 female) aged between 20 and 70 (mean age $\bar{x} = 36$ years, $\sigma = 16$ years) with varying degrees of skiing experience. After a brief introduction to the topic, the participants gave us information on their personal skiing preferences, skiing behavior and injuries. Subsequently, 21 cropped ten-second skiing video sequences in 1080p resolution and in randomized order from the experiment were shown to the participants. The volunteers were also informed to pay attention to the gliding sounds of the skis. After each video they had to select the corresponding slope conditions: wet, ice, groomed, packed, flat, mogul or powder snow. They could select and rank multiple snow types for each video. In addition, a brief textual and pictorial description was given for each snow type. They could watch each video as often as they wanted. Lastly, thoughts on the prototype were collected. We wanted to find out if the participants would use such a system. Furthermore, we asked questions regarding data privacy concerns and what details they payed attention to during the video labeling task.

6.5 Results Cycling

In this section, we present the results of our cycling study.

6.5.1 Hyperparameter Optimization

To narrow down the optimization space, the experimenter additionally recorded 34 minutes of cycling data including 20 % asphalt, 17 % cobblestone, 21 % gravel, 19 % paved, and 23 % unpaved paths. This additional data was used to evaluate the sliding window size, preselect classifiers, and preprocessing methods that were further evaluated on the study data. We tested the raw data stream and prepocessing methods from related work, including Fast Fourier Transformation (FFT) [Zür19] and statistical features (mean value, variance, standard deviation, first quartile, median, third quartile) [Mit19].

Figure 6.3 left shows the classification accuracy of the five recorded road surfaces obtained with statistical and FFT features on random forests (RF), support vector machines (SVM) [Bui13], and simple neural networks with dense layers [Cho15]. Talos and class-balancing was applied to optimize the network structure [Aut]. A sliding-window duration of 1.5 s yielded the most robust results on average. Hence, we further tested (shown in Figure 6.3 right) all preprocessing methods with a window size of 1.5 s using 3300 non-overlapping samples. 1D convolutional neural networks (1D-CNNs) achieved on average the best results followed by SVMs using statistical

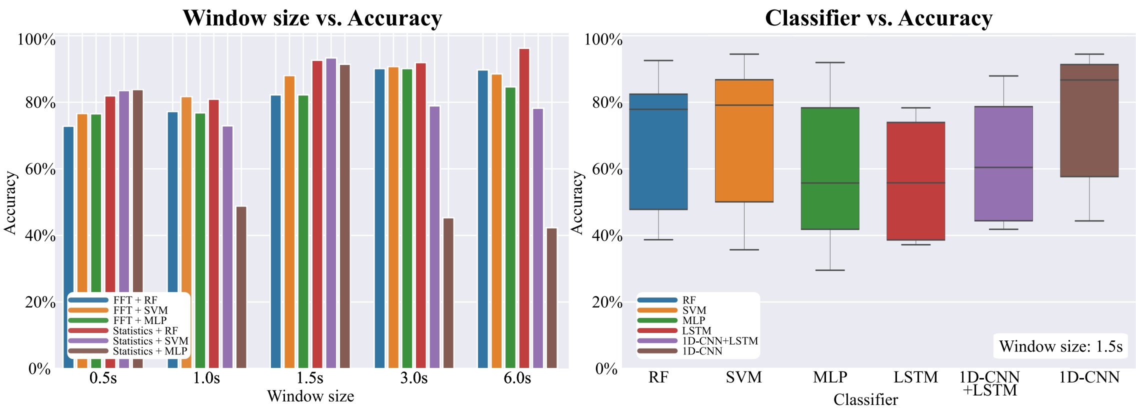

Figure 6.3: Window size and classifier evaluation. On the left: Performance comparison of various classifiers using statistical and FFT features. Right: Average performance comparison of different classifiers using a window size of 1.5 s.

features (92 %). 1D-CNNs fed with raw motion data reached 92 %, while audio alone reached 47.2 %. Concatenating both trained 1D-CNN structures by deleting the output layer and adding two Dense layers improved the accuracy to 94.5 %. In general, audio alone is not sufficient to achieve reliable results, but improves the performance by about 2 % compared to motion alone. Unwanted noises can propagate through the frame and affect the audio measurements.

6.5.2 Evaluation

For the entire dataset we use 1D-CNNs fed with raw data and SVMs using statistical features. Our study dataset consists of 42100 non-overlapping 1.5 s samples including 24 % asphalt, 7 % cobblestone, 26 % gravel, 27 % paved, and 16 % unpaved paths. We use the first twelve participants for training and the remaining five participants as our test set. On our training set we applied cross-validation for each participant. Figure 6.4, left, shows that 1D-CNNs with raw data achieve a better performance ($\bar{x} = 71.2\%, \sigma = 9.5\%$) than SVMs with statistical features ($\bar{x} = 64.1\%, \sigma = 7.9\%$). Participant 6 used a racing bike with a tire pressure of 6 bar (600 kPa), and P11 as well as P12 had a very low tire pressure of 1 bar (100 kPa). To confirm the assumption that the tire pressure influences the recognition rate, the cross-validation test was repeated on the 1D-CNN and the respective tire pressures of the participants were tested separately. Figure 6.4, right, shows that the optimal pressure for recognizing surfaces is between 1 and 4 bar (100-400 kPa). We conclude that tire pressure affects frame vibrations and consequently the recognition rate.

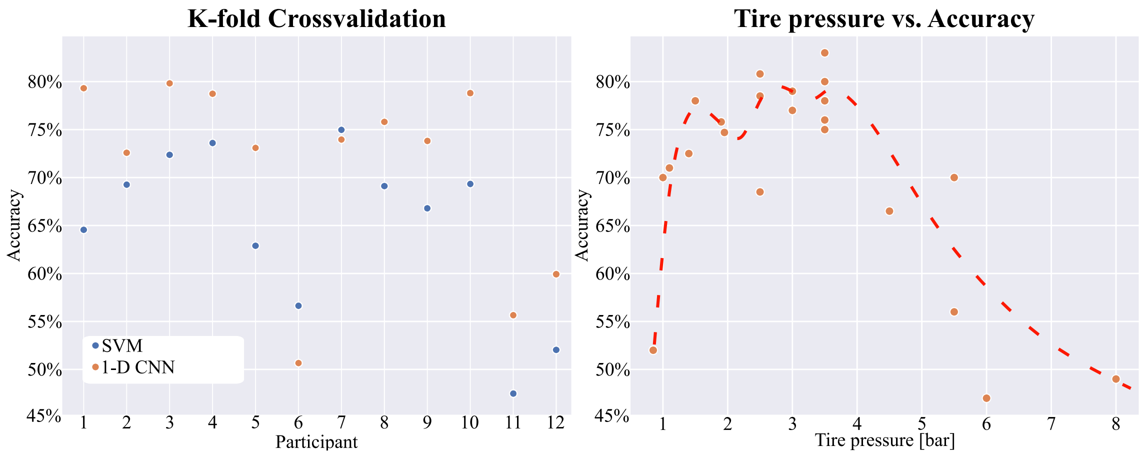

Figure 6.4: K-fold cross validation on the training dataset (left) and tire pressure in relation to the test set accuracy (right): The average tire pressure in the dataset ranges between 1.5 and 3.5 bar. Tire pressures for participants 6, 11, and 12 were out of that interval and achieved lower recognition rates. The red dashed line visualizes the influence of the tire pressure on the training dataset.

The test set (P13 to P17) includes four city bikes and one dutch bike with tire pressures ranging between 1.5 and 3.0 bar (150-300 kPa). The distribution of the cyclists' age and weight in the training set (age: M=30.5, SD=10.4 years; weight: M=74.9, SD=13.2 kg) to the test set (age: M=32.3, SD=14.3 years; weight: M=75.8, SD=11.8 kg) is comparable. Besides general tests, we evaluated the classifiers by using 10 % of the data of each test user for retraining 10 epochs. Figure 6.5, left, shows an average improvement of about ($\bar{x} = 8.2\%, \sigma = 2.5\%$) when individualizing the classifiers. In comparison to our DCI clustering benchmark [Wag20], which reaches 73 % on our dataset, we achieve similar results (M=75.3%, SD=5.2%). However, our approach offers the opportunity to retrain the classifier and to average across consecutive 1.5 s windows while DCI clusters are estimated over full road segments. By averaging results of three samples, or about 5 seconds of cycling, we achieve an improvement of 5 % without any changes to our 1D-CNN classifier. The dashed lines compare the average results of DCI clustering (red), generalized 1D-CNN combining audio and motion data (blue), and individualized 1D-CNN (green). As before, we identify an average improvement of about 2 % over DCI clustering when audio is combined with motion data. Again, we can conclude that tire rolling noise contains features which contribute to recognizing different road surface materials. The network structure is shown in Figure 6.5, right. The dense and CNN layers use ReLu activation and the output layer the softmax function. Audio and motion networks were first trained separately. Then, we deleted the

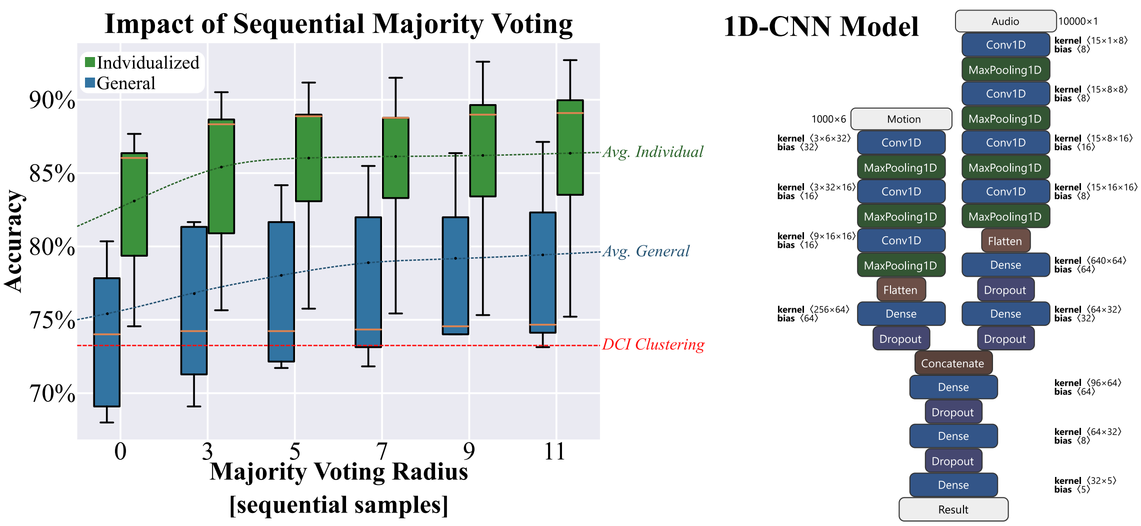

Figure 6.5: Left: Impact of averaging sequential classification results on the accuracy with and without individualizing the classifier on the test dataset compared to DCI clustering for benchmarking. Right: The model structure for combining audio and motion data.

output layers, concatenated both sub networks, and trained the entire network on the same data again. This strategy allows the classifier to identify important sensor features of the respective sensor.

6.5.3 Qualitative results

In the closing questionnaire, we surveyed our 17 participants about their bike usage, whether they have concerns to share cycling data and how road surfaces influence their choice of cycling routes.

Figure 6.6 gives an insight into the results of our questionnaire. In the first part on the left, we asked questions regarding our sensor kit mounted on the seat stay and smartphones or sensors mounted on the handlebars. Our respondents agreed that theft is a major challenge, whereas permanently mounted systems on the seat stay are inconspicuous and less likely to be stolen. Only one respondent stated a stolen part there while seven reported stolen parts on the handlebars. Three participants mentioned that the Cyclables had influenced their cycling behavior, but this was mainly related to the smartphone mounted on the handlebars. The mounting of a smartphone on the handlebars was considered cumbersome for daily tours. The respondents were concerned about forgetting their smartphone on the handlebars and would therefore prefer a permanently mounted sensor system for everyday rides. With this in mind, it was important to ascertain whether there are concerns about sharing the measured cycling data. In general, we obtained a quite positive sentiment from the respondents regarding data sharing, whereby only one participant stated to use the data for own purposes. However, the respondents had different opinions about the sensors. Motion data raised the least concerns, followed by audio data, and location data. To our surprise, necessary location data was judged as rather neutral regarding privacy. In contrast, sharing videos of road surfaces raised strong privacy concerns.

The second part of the questionnaire on the right of Figure 6.6 dealt with possible application scenarios in which the road conditions and sensor kits could play a role. Here, the everyday use of bicycles was of particular interest. The first set of questions addressed route planning activities. Here, the respondents stated that they usually have good conditions on their daily urban routes, which they mainly spontaneously adapt to the weather conditions. This means that when the respondents suddenly encounter impassable route segments on their daily tours, they try to cycle on known alternative paths. Navigation software is usually only used in unknown terrain. Poor road conditions can lead to increased bike maintenance requirements. Our respondents stated that they mostly treat their bikes carefully and repair their bikes on their own. Only for difficult repair cases, they consult a repair shop. In sporting activities, road conditions have a specific role. For recreational cycling and long distances, the most comfortable routes are selected, while for mountain biking, challenging trails are considered. However, weather conditions such as rain, snow, and wind are seen as additional challenges, and for this reason, respondents stated that they do not ride their bikes for sporting activities in all weathers. By comparing their own distances with the distances of other people, our surveyed cyclists were still motivated to become active. By reaching a specific goal, our respondents reported to be further motivated.

In the last part, we instructed our participants to rank preferred cycle road surfaces under different weather conditions. All votes are visualized as confusion matrices in Figure 6.6 on the bottom. Among all surveyed weather conditions, unpaved roads and cobblestone were rated as most inappropriate for cycling. Asphalt was remarked as the most preferred road condition in sunny weather, while our respondents ranked gravel and paved roads almost equally. On rainy tours, the picture changes and paved roads are rated as an almost equal alternative choice to asphalt. The most diverse opinions were given with snow on the roads. Asphalt and paved roads were ranked equally but gravel was ranked as a suitable alternative. In cold weather with black ice on the roads our respondents ranked paved paths slightly higher among the others. Gravel roads and cobblestone were rated almost equally, but similar clusters are formed as in sunny weather. This indicates that gravel roads tend to be ranked higher in this case.

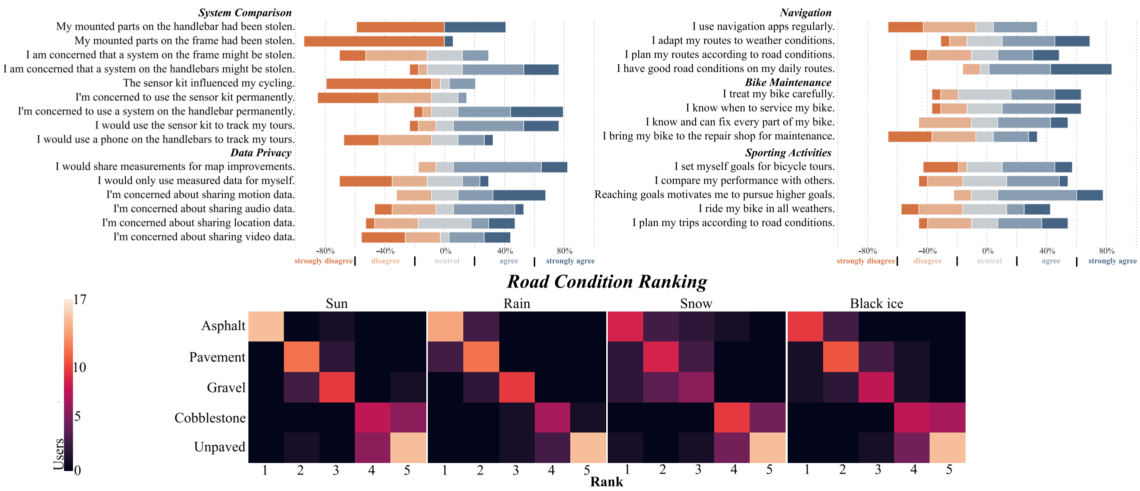

Figure 6.6: Results of the closing questionnaire. The road surface preferences strongly depend on weather conditions.

6.6 Results Skiing

In total, 3.18 GB of recorded data were used for the evaluation of our skiing experiment. Through variance in the connection stability, as expected with Bluetooth, the minimal achieved transmission rate was 576 Hz for motion and 9756 Hz for audio. Hence, we down-sampled motion data to 550 Hz and audio to 9500 Hz for further processing. We labeled the data manually with our notes and the recorded videos captured from the worn action camera acting as a reference. After discarding the irrelevant labels (e.g., idle or transit times and unreliable video data) we acquired in total 42:19 minutes of labeled snow segments that originate from eight descents.

The acquired segments were sliced into samples to form the training and validation set (with a ratio of 0.8 and 0.2 respectively). We first tested the final trained neural network from our cycling study. For this purpose, we retrained each sub-network for audio and motion on the skiing data and then combined the structure and retrained the network again. We tested the selected window size of 1.5 seconds as well as 1 second and 10 seconds. To match the different sampling rates we applied SciPy's Fourier resampling to the data [Vir20]. In no case was the network able to generalize the data. Hence we decided to create a new structure.

In total we have extracted 6027 samples with a window size of one second including 1640 'flat', 1434 'packed', 798 'wet', 658 'ice', 610 'mogul', 596 'powder' and 291 samples of 'groomed' snow. Our model consists of a one-dimensional convolutional input layer ('Conv1D') followed by two long-short term memory layers ('LSTM'). The Conv1D layer has 32 filters, a kernel size of 8 and a stride of 1. It uses the 'ReLU' activation function. The two LSTM layers have 64 and 32 units, respectively. The first LSTM layer returns its sequences to the second LSTM layer, before reaching the fully-connected output layer with 7 units. We used an Adam optimizer with a learning rate of 0.001 and trained the model for 1000 epochs.

Table 6.1: Accuracies (acc.) of the three trained models per data type. The best accuracy is the global optimum, whilst the last accuracy is the result after training for 1000 epochs.

ID	**Data set**	**Best acc.**	**Last acc.**
1	Audio and motion	0.54	0.50
2	Only audio	0.49	0.41
3	Only motion	0.44	0.36

Table 6.1 lists the highest and last accuracy that our model achieved per tested data set. We tested differed structures including convolutional layers, calculated spectral data, sensor measurements filtered by Butterworth low-pass filters, fast Fourier data with Hamming windows as well as deeper structures including regularization layers and dropout. The best performance was reached by training the model with raw

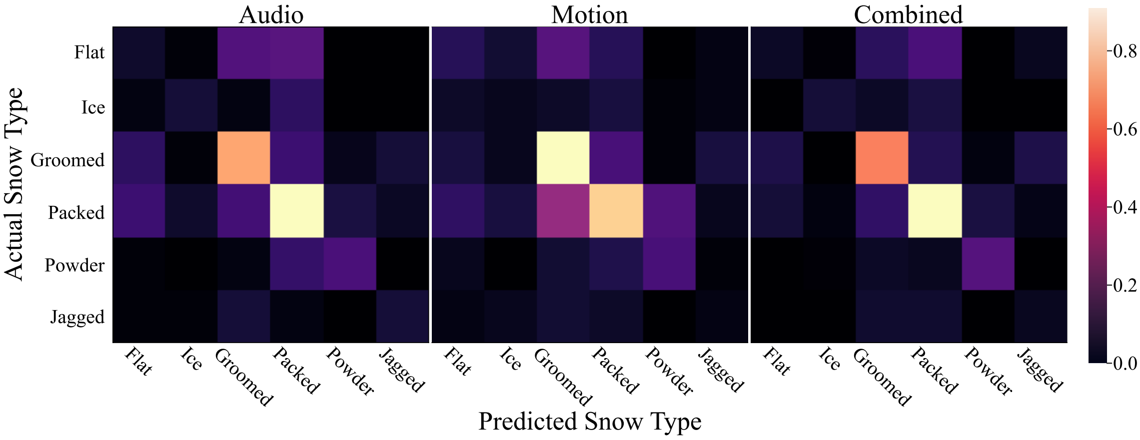

Figure 6.7: Confusions of the slope condition classifiers. The accuracy is not sufficient to identify the snow types.

audio and motion data. For this model, the highest tested accuracy was 0.54, whilst the last epoch's model resulted in an accuracy of 0.50. When training the same architecture only with audio data, the model scored a peak accuracy of 49 percent. Subsequently, when we trained the model architecture only with motion data, it reaches a highest accuracy of 44 percent. The corresponding confusion matrices are depicted in Figure 6.7. Generally, the preliminary results show that more data is required to reach meaningful recognition rates.

6.6.1 Online Survey

The 17 surveyed skiers ($age : M = 36{,}3\ years, SD = 16{,}3\ years$) skied on average 84 days in their lifetime with a standard deviation of 61 days. The least experienced skier stated 9 days and the most experienced 200 days, which gives us a wide range of experience levels. Nine of the surveyed skiers mentioned that they had participated in a skiing course. Five of them took part in a course to improve their technique, 3 took part in a school course and one gained his first experiences with an instructor. The majority, consisting of 8 skiers, stated to prefer red slopes, while 2 usually ski on unprepared tracks, 3 prefer blue slopes and 1 prefers black slopes. A total of 7 skiers reported to have injured themselves once or several times during winter sports activities. Two of them had to consult a doctor and stop their skiing trip. They mentioned that their injury was caused by an accident on a slope in poor conditions at high speed. In total, 5 of the skiers remembered that they had been skiing on an icy slope and 2 had an injury caused by moguls. Furthermore, 4 of the surveyed skiers commented that they had minor injuries on a slope with powder snow. Furthermore, 6 people stated that the accident could have been avoided by better slope conditions. In a comment field we collected further details about their injuries. We found that besides high speed, inattention was also a reason for their injuries.

Figure 6.8 gives an overview of our survey results. In general, our respondents stated that slope conditions have a strong influence on the winter sport experience. In addition to the route selection, the slope condition also influences the skiing behavior and the risk of crashes. Although maps of winter resorts have been rated as helpful, more detailed information would assist in better assessing a slope. Our respondents could imagine wearing a sensor kit, whereby motion data raised the least privacy concerns, followed by pressure data of the skis. About 40 % were concerned that audio could contain voice recordings and rated audio measurements lower than motion and pressure data. When no voices can be identified, our respondents were rather neutral about sharing audios. Sharing anonymized video was rated with less privacy concerns than audios. Although our respondents clearly remarked possible advantages of a sensor system, they had rather diverse opinions on measuring the slopes for tracking their experience. This may be due to the fact that the majority of people do not use digital tools to track their winter sports experience.

On the right of Figure 6.8 the ratings of the weather conditions are visualized. Sunny and cloudy weather are accordingly the most suitable conditions. Under more challenging conditions, our respondents were not unanimous in their assessment. In case of bad weather winter sports are partly not possible. Therefore, no clear clusters can be formed. The picture is different for the snow conditions on the slopes where two bigger clusters can be derived. Flat, groomed and packet snow form a cluster as rather good conditions while wet, jagged and icy slopes were remarked as challenging.

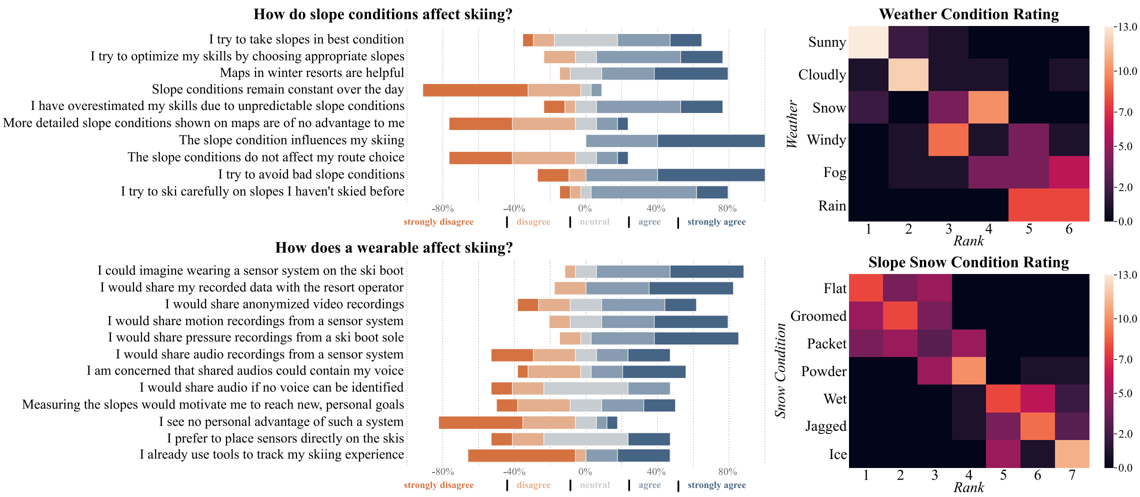

Figure 6.8: Questions regarding skiing preferences and wearables. The surveyed skiers agree that the slope conditions influence their skiing style and route selection. They could imagine using a wearable for their experience, if privacy issues are considered.

Powder snow was rated in between both bigger clusters with a tendency to form a small cluster with packet snow.

During the online survey, participants had to perform a labeling task to validate our labels noted down during the experiment. The results and example pictures from the video sequences are visualized in Figure 6.9. We had a rather diverse labeling performance among our respondents. They stated to use audio as well as video sequences for their decisions but found it rather difficult to imagine a label. In contrast, our respondents were confident that they could identify all conditions when skiing on the slopes. They remarked that they mostly looked at grooms, whirled up snow or camera juddering. Scratching sounds were identified as icy slopes, while low noise levels were classified as powder snow. The changing visual conditions were perceived as challenging by the participants. When videos gave them no clear identification, they used audio features to make their decisions.

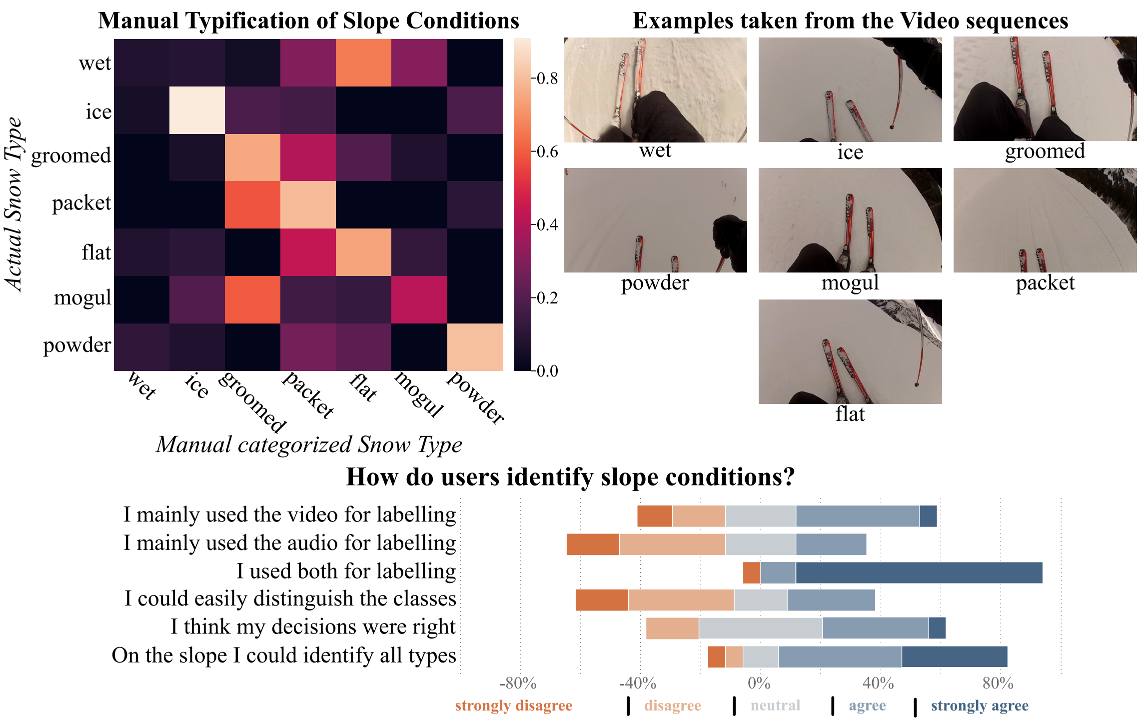

Figure 6.9: Questions and results of the labeling task. The participants used both the video and audio material for their decision. They were confident to be able to distinguish the different classes, if they did the ride themselves. All of them agreed that the different conditions influence their skiing style.

6.7 Discussion

The conditions of a surface are important in outdoor sports activities. They have an influence on the passability of a route and the involved risk of injury. In the wild, the conditions are dynamic and weather dependent. This also influences the optimal route selection. Our results show that especially in challenging weather conditions during the cold season, cyclists also consider alternative routes on gravel paths as suitable alternatives. By visualizing the ratings as matrices, weather dependent clusters can be derived. In all weathers, cobblestone and unpaved roads are the least preferred conditions. For winter sports activities, an even greater dependence on the weather can be observed. Here, only cloudy and sunny weather conditions were preferred. Under more challenging weather conditions, our participants were not in agreement about a ranking. This indicates that they tend to avoid winter sports under these conditions. With regard to the slope conditions, two clear clusters were formed. Our respondents rated flat, groomed and packet snow as good quality slopes while wet, jagged and icy slopes were rated as bad conditions. They mostly agreed that ice involves the most challenges during the activity. All slope conditions are dynamic and independent of the static categorization, which shows that a more precise specification could support route selection tasks. Similarly, an integration into route planning tasks for cyclists could help suggest alternative routes. Especially novel forms of multi-modal mobility on the last mile are associated with higher risks of injury and could benefit from the entered information.

To measure surfaces in the wild, we proposed to use audio and motion measurements. While using motion sensors for road roughness measurements and road surface classification on cycled paths is an established method, audio data were so far used for environmental noise sensing. Our results show that the sounds produced by tires contain information that can be used to enhance the recognition of the road surfaces. Audio alone cannot replace roughness measurements, but increase the recognition rate by at least 2 % when combined with vibration data. To incorporate the individual sensor features into the classification, we first trained each sensor sub-network and then combined the 1D-CNNs by deleting the output layer and adding dense layers. This final structure was then trained again on the training set. By using this final network we identified that the tire pressure influences the produced sounds and vibrations. As a result, different tire pressures affected the recognition rate. However, within most of the tire pressure range from our dataset, the proposed network structure achieves a relatively constant recognition rate. A particularly low as well as high tire pressure resulted in a strong influence. It can be deduced that an adaptation of the structure to a given setup is necessary. Individualization of the 1D-CNN network by retraining with user-specific data resulted in an increase of the recognition rate by 10 % on average. Averaging successive 1.5 s samples resulted in an additional average increase of 5 %. Since maps partly contain information about the road surfaces, available data could be used to automatically individualize the classifier to a given setup. In the wild, further improvements could be achieved by averaging the classification results on a route segment across all measured trips. To compare our approach we used clustered road roughness DCI measurements [Wag20]. The improvements over clustered DCI-measurements [Wag20] of 2 % are comparatively small at first glance. However, our approach has a much finer granularity, as it requires only 1.5 s of data whereas DCI-clusters are classified over an entire road segment. Apart from the possibility of increasing accuracy by averaging over adjacent measurements, our approach is less prone to errors when cycling over curbs, since only single classification results are affected. Especially on short road segments, DCI measurements react strongly to data peaks. A hybrid approach that validates the DCI clusters with our 1D-CNN would be able to filter out such anomalies. Nevertheless, such peaks are indeed of interest, because they represent important road surface quality parameters, such as the presence of potholes. However, cyclists usually circumvent potholes when possible. Moreover, as curbs and potholes affect single classification results, our approach filters out such anomalies and focuses on the general surface condition of a road.

Adapting the proposed 1D-CNN structure for skiing was not successful and even an optimized structure based on 1D-CNNs followed by LSTMs could not reach acceptable recognition rates. This can be related to several reasons. First, the amount of data is limited as we collected a relatively small dataset of a half day of skiing. Second, the conditions on the slopes are highly dynamic. A skier can have optimal conditions in the middle of a slope, but encounter snow bumps or powder snow at the edges. In addition, the conditions can change very dynamically. This makes it more challenging to correctly label collected data from the slopes. Our subsequent online survey could confirm this assumption by the results of the labeling task. A few classes were identified with certainty from audio and video data. Icy slopes were the most accurate conditions that could be manually identified by the videos. This can be related to the scratching sounds. Powder snow can be more easy identified by whirled up snow. Fresh prepared slopes are easy to identify by the grooms in the snow. In contrast, snow bumps are challenging because they are not clearly visible in the camera image. The same applies to old (wet) slopes. The partly melting snow is difficult to recognize in the videos. Four of our respondents also remarked in a free comment field at the end of the survey that the labeling task was not easy. One respondent stated that the image contrast of snow is not sufficient to categorize the slope conditions. It can thus be argued that the slope conditions should first be investigated by using artificially created environments as in a ski hall. Based on these results, a field study could then be conducted. It would be of interest whether displaying additional slope conditions could influence skiers' route choices. In a winter resort, it could then be investigated whether the independent variable display of slope conditions in maps has an influence on the dependent variable number of falls.

In general, it can be deducted that there is a weather dependent relationship between preferred surface conditions and route choices. The noise produced by frictions on a surface is not sufficient to reliably classify the surface conditions. Combined with motion measurements, slightly higher recognition rates then by using only motion can be achieved. If a sensor kit already collects sound data, then these measurements can be integrated into the road surface classifier to improve the recognition rate. Measured audio data raise more privacy concerns than motion data. Highly dynamic changes of surface conditions, like on skiing slopes, are challenging to distinguish. It is necessary to collect data in a controlled setting before an approach in the wild can be investigated. In contrast, surface conditions on cycled roads in urban areas are sufficiently constant to collect and analyze data.

To provide classified road surfaces for cyclists the official map of *OpenCycleMap* [Ope20] can be used. For this purpose, the measurement data or the classification results from the user's smartphone must first be stored on a server. These data are then averaged over several trips on a particular untagged road segment and entered into the public available maps. If an individualized classifier is applied, the measurements of tagged route segments must be sent to the server. The server then individualizes the 1D-CNN to a specific setup based on the labels and sends the resulting classifier back to the smartphone. Measurements can be temporarily stored on the user's smartphone and sent to the server when a WiFi connection is available. A similar method could be implemented in winter resorts. Here, the ski lifts or stations would be equipped with WiFi. During lift transportation or waiting for the lift, data would be sent to the server. The map including current slope conditions and traffic could then be displayed on screens at the stations, in the lift or via an app. To collect a large amount of data, ski rental shops could equip their shoes with the sensor. A discount on the equipment could convince skiers to use the system. One challenge is that users need to daily charge the batteries. Therefore, the discount could be depended on the amount of shared data. Furthermore, ski instructors could mount the system on their boots.

6.8 Conclusion

This chapter investigated how surface conditions in the wild can be tracked with audio and motion data. For this purpose, we have built and tested a sensor kit in two application scenarios. At first, we evaluated our sensor kit for cycling and then tested the approach in a skiing experiment.

For the cycling study, we recruited 17 volunteers to collect a dataset including 130 km of cycling on different bikes and common surfaces in urban areas. We showed that audio, which was up to now mostly used for environmental noise monitoring, can slightly improve the recognition of common surfaces: asphalt, cobblestone, gravel, paved, and unpaved paths. One-dimensional convolutional neural networks that are fed with the raw data, achieved an average recognition rate of 88% after retraining the classifier on user-specific data and averaging consecutive classification results. Without individualization, we achieved a recognition rate of about 75%. This reduction in classification performance is not surprising, as there is a wide variety of bike technology (frame stiffness, tire pressure), user properties (body weight), and cycling styles (speed, gearshift points). As maps partially contain road surface information, data that is collected while cycling on already tagged paths can be used for individualizing a classifier, without the user needing to be aware of this continuous training process. We compared our approach to commonly used DCI clusters that were computed from smartphone IMU sensor data. We found that our proposed system is able to make predictions with a finer granularity than when using DCI clusters, i.e., it can handle shorter road profile segments. However, individual characteristics related to the user, the cycling behavior, tire pressure, and the used bike were found to strongly influence the accuracies of the classification results.

Furthermore, we found that route preferences of cyclists depend on the weather conditions. Gravel roads were rated as a suitable alternative in the winter season. Although gravel increases the probability of soiled clothing, our respondents rated it higher than cobblestone. However, the route planning of current navigation systems does not fully consider preferred surface types depending on weather conditions. Commonly used static road roughness measurements are not sufficient to categorize a road, since weather-dependent properties cannot be derived from vibration measurements alone. The weather-dependent preferences of road surface types must therefore be taken into account in route planning. The Cyclables sensor kit allows estimating road surfaces at higher accuracies than DCI clustering approaches. Subsequent automatic entry of classified road surfaces into maps enables to adapt the route based on the current weather conditions. Parameters such as soiled clothing or safety in icy conditions can then be taken into account based on the weather report. Privacy concerns influence the willingness to share data. Since location data are necessary to link measurements to a certain position, sensors such as cameras, which are more concerning, may cause a decreased willingness to share data. We conclude that sensor kits that already contain audio sensors, e.g. for ambient noise monitoring, benefit from combining audio with motion data if the road surface conditions are to be recorded by vibrations. In order to increase the willingness to share data, we surveyed the bike usage of our participants. It turns out that sensor kits should offer additional incentives to support cyclists on their daily trips. Those incentives could be related to theft, sporting activities, and bike maintenance. Navigation software for cyclists is rarely used in everyday life. Therefore, mounting a smartphone on the handlebar for every ride can be perceived as inconvenient. Like car drivers, cyclists mostly use navigation software in unknown territory, e.g. when the known daily route is closed due to road works or for sporting and recreational activities.

The subsequent skiing experiment was conducted to evaluate our sensor kit in a different scenario. We mounted the system on a ski boot to measure slope conditions in a winter resort. The aim was to establish more detailed maps in winter resorts that also take into account slope conditions to increase safety and improve the sporting experience. From a half day of skiing, data from one experienced skier were collected. We found that slope conditions in the wild have a highly dynamic fashion. A slope may have good gliding properties in the center, but contain snow bumps or power snow at the edges. Conditions may also change temporarily on a segment, making it difficult to accurately label the data. As a result and due to the limited amount of data, our proposed 1D-CNN from the cycling study could not reach meaningful accuracies. A second network structure was proposed and optimized based on 1D-CNNs and LSTMs where we could reach accuracies of about 50 %. Since the results are not sufficient to extend maps in winter resorts, we assume that data must be first measured in a controlled setting e.g. ski hall. This was also confirmed in our online survey where we evaluated whether a labeling task could be performed on the basis of video data. In total, 17 participants labeled 21 video sequences from the skiing experiment with a length of 10 s each according to their slope condition. We found that snow offers insufficient contrast to be accurately identified from videos. In addition to the video recordings, the corresponding sound is also important, as scratching sounds indicate the presence of ice. Further questions about preferred slope conditions showed that skiers avoid bad weather conditions. Regarding the slope conditions, our respondents rated wet, jagged and icy slopes in a cluster as poor slope conditions while flat, groomed and packed snow were rated in a cluster as good conditions. A slope may be categorized as easy but may involve unpredictable risks due to ice. This showed that a static categorization is not sufficient. An additional dynamic rating that simply discriminates good or bad slopes could already support the route selection.

In general, it has been shown that surface conditions are related to the current weather conditions. Since route preferences are also weather-dependent, a static categorization is insufficient. As a result, measured and categorized vibrations on a surface alone are not meaningful enough. Scratching sounds produced by the friction on the surface combined with vibration measurements contribute to a more accurate classification of road types. There are various scenarios where a road surface estimation system would be helpful. However, the skiing experiment showed that labeling can be difficult due to the very dynamic changes of the slope conditions. The surface preference clusters revealed that a categorization between good and bad conditions is already sufficient to support route selection. In the case of cycling, publicly available road surface map data already exists that can be used to autonomously adapting the classifiers to a given bike setup without involvement of a user. In the future, our proposed system could be extended to novel forms of multi-modal mobility on the last mile, e.g. e-scooters.

CHAPTER 7

Using Passive Surfaces for Smartphone Interactions

The smartphone has become a regular companion in everyday life. Even during work, the smartphone has its place on the office desk for many people to quickly respond to incoming messages. This chapter deals with how surfaces can be utilized to identify the location of a smartphone and how sensed surfaces can support smartphone users. For this purpose, an approach is presented that recognizes different materials based on color features of a surface. We use the rear camera of a phone with the LED flashlight to identify specific characteristics related to color and reflection of a surface. The results are then compared to a previous approach that used the front camera and different display illuminations [Yeo17] for surface sensing. Since our approach offers the possibility of keeping the display visible for the user, we will subsequently motivate future applications. Thus, this chapter will address the following research questions.

> Can a phone's rear camera and white LED flashlights be utilized to identify surfaces?
>
> Are there more interaction capabilities if the display remains visible?

This chapter is based on the master thesis by Philipp Etgeton [Etg18], a semester project [Sad22] and a late breaking work at *CHI 2021*. In his thesis, Philipp has built and tested a prototype to analyze different external flashlight LEDs. Based on the results which were published as late breaking work [Sch21], the semester project served to implement and test the technique with internal flashlight LEDs for notification management. At the same time, Maximilian Schrapel implemented an example application to locate a smartphone by surface sensing.

- Schrapel, Maximilian & Etgeton, Philipp & Rohs, Michael: 'SpectroPhone: Enabling Material Surface Sensing with Rear Camera and Flashlight LEDs'. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021

7.1 Introduction

Today, mobile phones are highly integrated systems with a variety of sensors for different applications. For instance, inertial measurement units can track the position of the phone for applications such as bubble levels and games, gestures, or pedometers. In addition, rear cameras and flash light LEDs can not only illuminate a scene but also measure the heartbeat [Pel10], blood pressure [Wan18] and Hemoglobin [Wan17] through the skin of a finger. Those and many more applications indicate that the phone is a ubiquitous multi-functional toolbox [Bal06].

Previous research has shown that a front camera in combination with different display color illuminations allows to capture spectroscopic features of surfaces [Yeo17]. During the sensing process, the display can not be used for other purposes, e.g. to give feedback on the progress. In this chapter, we show that the rear camera and white flashlight LEDs are sufficient to capture spectroscopic features. Since white light covers the entire visible spectrum, the number of captured images can be reduced. Due to API limitations, we use external LEDs mounted in a 3D-printed phone case to test different illumination levels.

Based on the proposed technology, we motivate two application scenarios where surfaces can support daily tasks. First, we present an approach to manage notifications in different scenarios. Twitter is used to illustrate how users can be presented with information that is related to their context. When the phone is placed on the working table only messages are displayed related to their office work. These can be, for example, twitter messages from colleagues or call for papers from important conferences. When the phone is placed on the coffee table, only messages related to lazy activities and friends are presented. Therefore, work-related tweets will not be displayed. The goal is a more efficient use of social media in an everyday context. Second, we motivate to integrate surfaces into the search for a personal smartphone in an home or office environment. When a user triggers tones on a smartwatch to search for their phone, additional information about the smartphone's location and surface is displayed on the smartwatch. This aims to make the search of a smartphone more efficient. Particularly in the case of hearing impairment, users can benefit from such interaction technology, as selective hearing is possibly lost and sounds can no longer fully assist search for the phone.

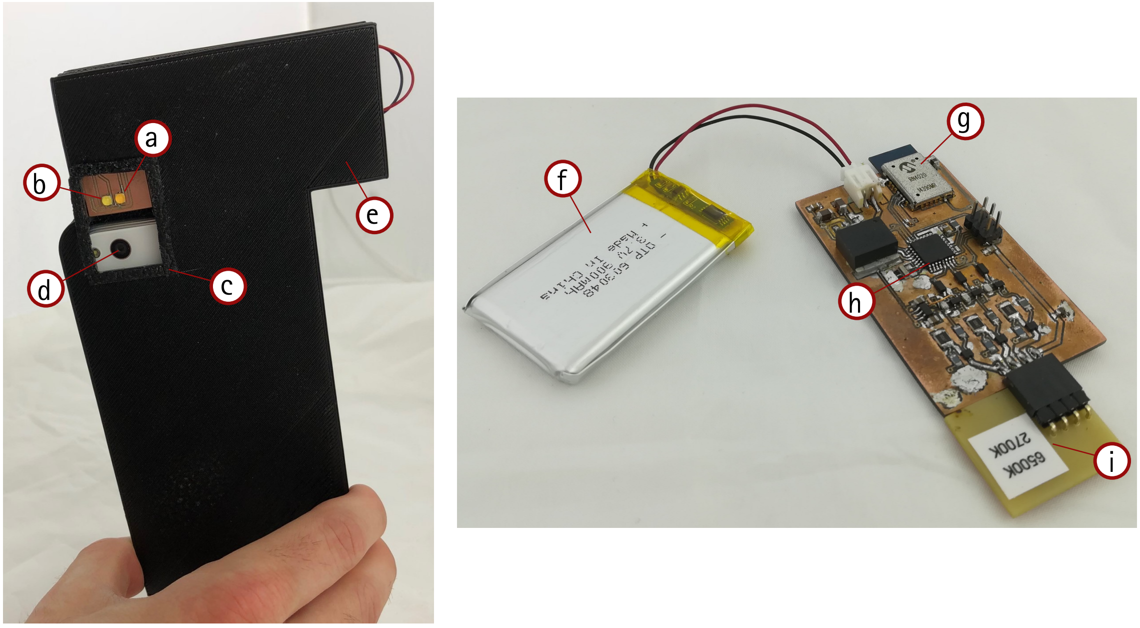

Figure 7.1: Rear view of the prototype. The case of the used smartphone (Huawei P10 lite) contains an extra control unit for the warm (see marker a) and cool white (marker b) LED via Bluetooth. A small frame made of black foam around the sensing window (marker c) prevents ambient light from entering and creates a fixed distance of 3 mm to the target object (marker d). The black housing color has the lowest influence on the measured spectra and is 3D-printed from ABS material (marker e). The battery powered (marker f) PCB for controlling the LEDs (marker i) is shown in the left corner. Bluetooth commands (marker g) are processed by a ATmega microcontroller (marker h) to generate PWM signals.

7.2 Related Work

Material sensing with smartphones based on spectroscopic features has been performed in the past with different methods. The methods can be divided into approaches that require additional hardware [Har08a; Sat15; Wan16b] and those that use only built-in capabilities [Sei; Yeo17] or even embed a molecular spectrometer [Cha17b]. All rely on the same principle: Emitted light is reflected by a material and then captured by a sensor. The remaining light is spectrally decomposed, processed and compared to a dataset. Various light sources have been examined in the past: RGB LEDs, also combined with infrared and ultraviolet LEDs or lasers [Sat15]. Without any additional equipment the smartphone display can also be used as a light source and the front camera is then utilized as a sensor [Sei; Yeo17]. Moreover, front cameras and flashlights have been used together with additional equipment for colorimetric biosensing applications [Wan16b]. Additionally, the company Two-Photon research is exploring the opportunity of using smartphone front cameras for instant COVID-19 tests [Res20].

In addition to cameras, inexpensive light sensors can also detect spectral components of a material [Har08a]. It has been found that expensive spectrometers do not provide good results in comparison to mobile applications because they measure the light only at a small spot and do not detect any variations in the surface texture [Yeo17]. Dedicated sensing equipment has been used for robots to sense object's fragility [Eri19]. Besides spectroscopic material sensing there has also been research on measuring object resonances with vibrations [Dem20; Oh19] and varying slicing parameters to distinguish different 3D printed objects [Dog20].

In order to distinguish from related work, our goal is to explore whether flashlight LEDs and rear cameras can achieve comparable results to SpeCam [Yeo17]. Keeping the display visible would allow further interactions like showing the sensing progress or giving feedback on interfering light. Due to restrictions in controlling the intensity of the flashlight LEDs of mobile phones [App18; Goo18] without hacking the operating system, we evaluated our approach with external white LEDs (similar to the built-in ones) at a range of intensity levels. In addition, we test our algorithm with built-in flashlight LEDs at full intensity.

7.3 Spectroscopy with white LEDs

Modern smartphones use dual or quad LED flashlights with different color temperatures to illuminate a scene. Their purpose is to achieve more natural colors in the resulting picture depending on the ambient light conditions. For example, in a relatively dark location where warm white light sources with a color temperature between 2000 and 4000 Kelvin are used, cool white LED flashlights with a color temperature of about 5500 to 6500 Kelvin produce an unnatural image. Therefore, warm white flashlights are applied to achieve an adequate illumination with lower blue and higher red components, as can be seen in Figure 7.2. Furthermore, modern smartphones use flashlights with a high color rendering index (CRI) of over 80 which describes the faithful reproduction of colors [lEc95].

Based on these fundamentals the full visible light spectrum can be reproduced by dual LED flashlights. In order to overcome differences in the hardware configuration of different manufacturers, a simple calibration on a white surface, e.g., a piece of paper, may be performed once. The spectral response can then be adapted to a comparative dataset.

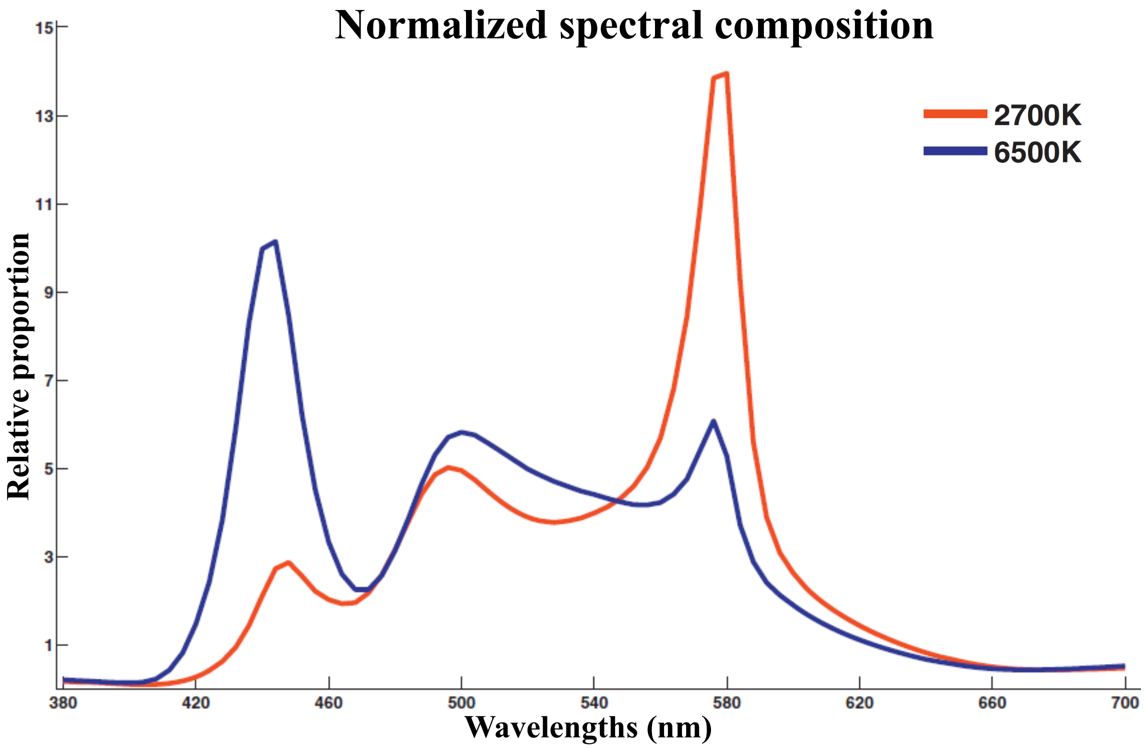

Figure 7.2: Normalized spectral composition for 2700 K (orange) and 6500 K (blue) light sources [Hue16]. It can be seen that cool white LEDs apply more spectral power on the blue component while warm white LEDs focus on the red component. For our studies we used two OSRAM Oslon LEDs with the same distributions.

7.4 Methodology

Our additional hardware unit inside the housing consists of a Bluetooth transmitter that is connected to a microcontroller that controls the external LEDs by pulse width modulation (PWM) via a current source. We use 700 mA OSRAM OSLON LEDs with a high CRI value of 90 to ensure accurate color reproduction. To avoid streaking (vertical stripes) in the resulting image, the PWM frequency must be significantly higher than the refresh rate of the smartphone's camera. The ATmega 8 microcontroller supports a PWM base frequency of up to 31.372 kHz which avoids any vertical streaks in the captured image. We also tested the default PWM frequency of 490 Hz which produced vertical streaks in the image from a camera with a refresh rate of 30 Hz.

The smartphone application sends commands via Bluetooth to the microcontroller where the PWM signals are generated to control the flashlight intensity. Each image is captured with the maximal resolution of 3986 x 2976 pixels with no auto-focus and ISO 80. For a full sample, 15 images have to be taken based on the LED intensity level. We control each LED in 20 % intensity steps resulting in 5 images per LED. Then we repeat the procedure by controlling both LEDs simultaneously, which results in 5 additional images. All pixels of each image are then broken down into their RGB components and divided into histograms with 64 bins per color channel as described in [Yeo17]. After a normalizing the histograms based on the maximum value, all histogram bins are concatenated to a feature vector. Per image 64×3 histogram bins are generated, which results in a maximum of 2880 feature components for 15 images. The resulting vector is then transmitted to a server to be classified by an Scikit-Learn Support Vector Machine in Python 3 [Ped11]. Our server is a conventional PC using Ubuntu OS for running a Python script to receive feature vectors, classify spectroscopic features, and send the classification result back to the smartphone.

7.5 Dataset & Study

In order to evaluate and compare our approach with related work, we have chosen 30 similar materials as those selected by Yeo et al. for SpeCam [Yeo17]. In summary for each sample per class 15 images were taken, as described in Section 7.4, resulting in 450 pictures for 30 samples. In total for 30 classes 13500 images are contributing to our dataset with a resolution of 12 megapixels. Each sample was taken at a different position and alignment on each material as described by [Yeo17]. The dataset together with the casing and PCB layout is available at GitHub [1].

We aimed to evaluate whether the rear camera of a smartphone for image capture and warm and cool white LEDs as the light source are able to achieve similar recognition rates as related work with the front camera for image capture and the front display as the light source [Sei; Yeo17]. For this purpose we chose three flashlight configurations: *2700 K* only, *6500 K* only, and *both* light sources together. For each configuration we evaluated different light modes: *single mode* with 100 % light intensity, *dual mode* with 20 % and 100 % intensity, and *full mode* with all contributing images per sample. When only one LED is controlled, 5 images are taken in *full mode*. With both LEDs in *full mode* 15 images are contributing to a sample. For our dataset all samples were taken in *full mode* with both LEDs. Other configurations can be derived by selecting the corresponding features. The goal of our tests is to achieve a trade-off between the number of images and the recognition rate while minimizing the time to capture. The maximum feature vector length when both LEDs and full mode are applied is 2880 components resulting from 64 histogram bins for 3 color channels and 15 images per sample. In single mode with one LED, a minimum length of 192 components is used. If both LEDs are controlled in dual mode, each single LED is controlled individually and then also together with 20 and 100 % intensity resulting in a feature vector with 1152 components.

1 https://github.com/M-Schrapel/SpectroPhone

The SVM used for classification is also examined more closely. It has to be determined whether the selected materials can be distinguished by means of a simple dividing hyperplane. Decisive for this are the kernel function and the size of the margin of the decision boundary, which is set via the penalty parameter C. We use 10-fold cross-validation with a split ratio of 70:30 to evaluate our models.

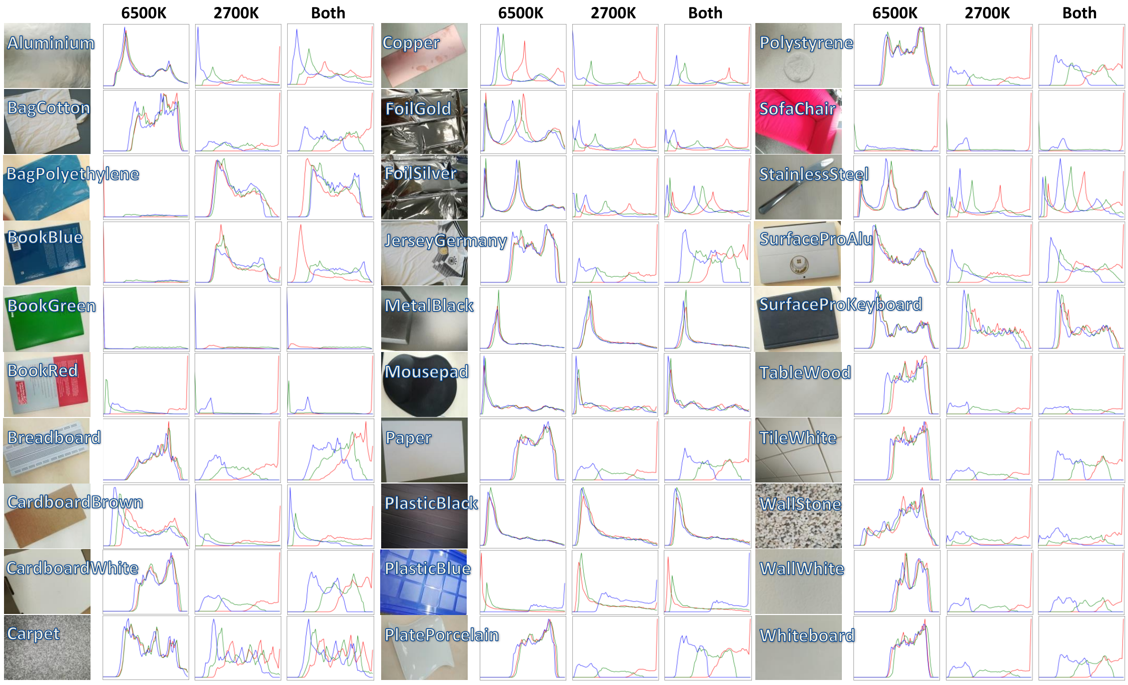

Figure 7.3: Subset of used materials and corresponding histograms with different LED configurations. The histograms differ clearly from each other. The classes *BookGreen* and *BookRed* show high proportions on another color at low color values due to the additive color mixing. It is noticeable that the resulting histograms of the warm white LED are more similar to the spectrum of both LEDs.

7.6 Results

To estimate the performance of our approach and find the most appropriate SVM parameters, GridSearchCV [Ped11] is applied on all LED configurations and SVMs using linear, rbf and polynomial kernels. We tested C parameters from 0.001 to 10000 for our analysis. The best results for each LED configuration and light mode are listed in Table 7.1.

In comparison to related work, white LED flashlights and the rear camera also provide high recognition rates of about 99 % which is similar to the results of SpeCam [Yeo17]. The linear and rbf SVM kernels show the best results for the most configurations. Beside of the entries in Table 7.1, where the rbf kernel had the highest accuracy, the linear kernel always was close, as exemplified for both

Table 7.1: Predictive accuracy of the best results with a 10-fold cross-validation.

LED	Intensity mode	SVM Kernel	SVM C-Parameter	Accuracy	Standard deviation
6500K	single	rbf	1000	0.950	0.0181
6500K	dual	linear	60	0.948	0.0172
6500K	full	linear	60	0.951	0.0187
2700K	single	rbf	1000	0.987	0.0097
2700K	dual	linear	60	0.986	0.0122
2700K	full	linear	1	0.983	0.0090
both	single	rbf	60	0.989	0.0086
both	dual	linear	60	0.988	0.0105
both	single	polynominal	60	0.988	0.0092
both	full	rbf	10000	0.989	0.0099
both	full	linear	60	0.988	0.0105

LEDs in full mode. By further evaluating the applied SVM parameters found by GridSearchCV [Ped11] it should be mentioned that the chosen features tend to be linearly separable by a decision boundary. For instance the polynomial kernel has the highest recognition rate when the degree of the equation is set to 1, which is almost equivalent to the linear kernel. The best performance was achieved when both LEDs contribute one image at full light intensity to the feature vector. While SpeCam [Yeo17] uses 7 images, flashlight LEDs can reduce the time for material sensing since only two images are sufficient for comparable recognition rates.

A further evaluation of the confusion matrix shows which classes are confused with each other despite an optimal choice of parameters. For this purpose in Figure 7.4 the linear SVM with both LEDs and all 15 images per sample was chosen. The only remaining confusion relates to the black metal and black plastic surface. This is reasonable, because black maximally absorbs the light. By reviewing the spectral response of both materials in Figure 7.3 it can be seen that they produce similar spectral responses. The same applies to white surfaces like paper and porcelain. However, white paper and porcelain differ more strongly in their reflection which results in more accurate classification results. Likewise, the black mousepad had less reflections than the other black materials resulting in a more different spectral response.

From the results, it can be concluded that different materials can be distinguished with white LED flashlights. It is not essential to illuminate a material with 7 different colors with a limited amount of surfaces. In addition, single images and flashlight LEDs are also sufficient, since only the proportions of the wavelengths differ in the emitted light spectrum as demonstrated in Figure 7.2. A limitation is that we could not control the internal flashlights in different intensities to establish our results and a distance between the camera and the material is always necessary.

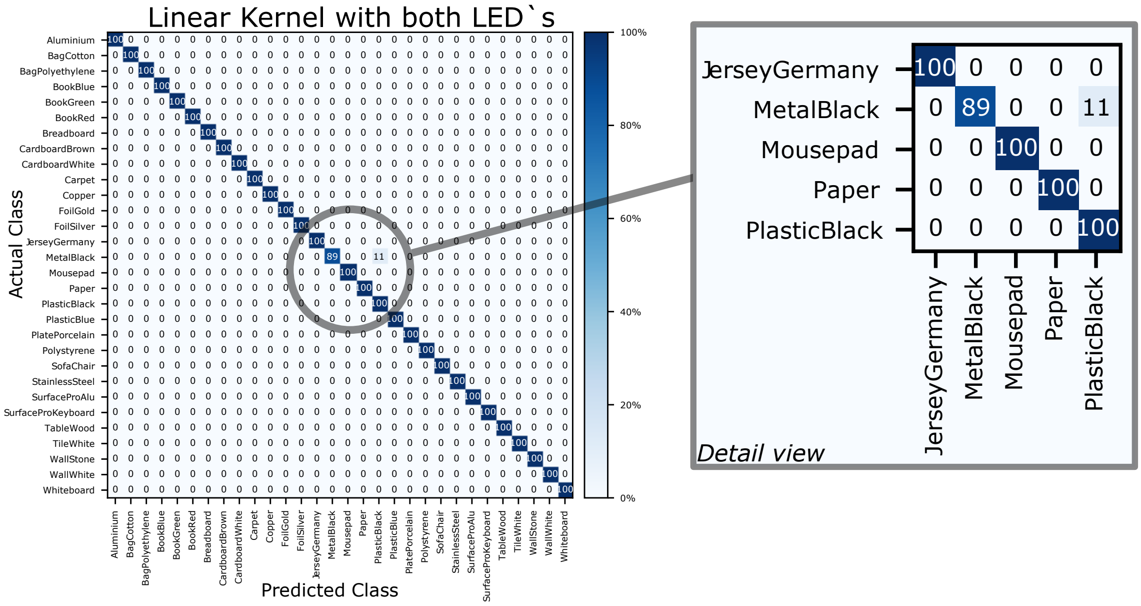

Figure 7.4: Confusion matrix for a linear SVM with both LEDs and all 15 images per sample. Only black surfaces show confusion, because for them the absorption of light is strongest.

7.7 Application Scenarios

In the following, applications are presented where the surface on which the smartphone is placed can support interactions. In contrast to the previous evaluation, only the internal flashlight LEDs are used and camera captures are automatically triggered. We evaluate the performance on several household materials and continue by presenting two applications. In the first scenario, notifications are managed by the surface. The idea is that, depending on the smartphone's location, the user performs different tasks. To exemplify, the phone is placed on the office desk and the user wants to be able to respond to incoming messages. The second scenario is based on the search for a phone in a home or office environment. By presenting additional information (e.g. surface type and color as well as an image of the smartphone front camera) we aim to make the search for the phone more efficient.

7.7.1 Smartphone Material Sensing without Additional Equipment

Before we present example applications, we evaluate the performance with internal flashlight LEDs. For this purpose, we attached a 3 mm thick neoprene shield on the back of a Xiaomi 9T pro smartphone to prevent external light from interfering the image captures. Materials from a home environment were selected based on the surfaces on which the experimenter usually places his smartphone. In addition, materials such as wood from a doorstep and carpets have been added. In total, 16 different household surfaces in a resolution of 720x1280 pixels have been collected.

For each material 30 images were captured and analyzed.

Figure 7.5 shows the resulting confusion matrix, materials and the smartphone with neoprene shield around the camera and flashlight. By using 10-fold cross validation, we tested the same SVM configurations as before with external flashlight LEDs. A Gaussian (rbf) kernel with a C parameter of $C = 1000$ showed the best results with an average accuracy of $M = 92.3\ \%$ and a standard deviation of $SD = 3.4\ \%$. Compared to Table 7.1, the same parameter set as with single flashlight LEDs were selected. However, the accuracy is lower, which can be mostly explained by looking at the materials and confusions in Figure 7.5. Especially the four black surfaces again show confusions among each other. The PVC floor closely resembles the color of the office desk resulting in higher confusions. It can be deduced that a heterogeneous distribution of colors is advantageous in order to distinguish the surfaces.

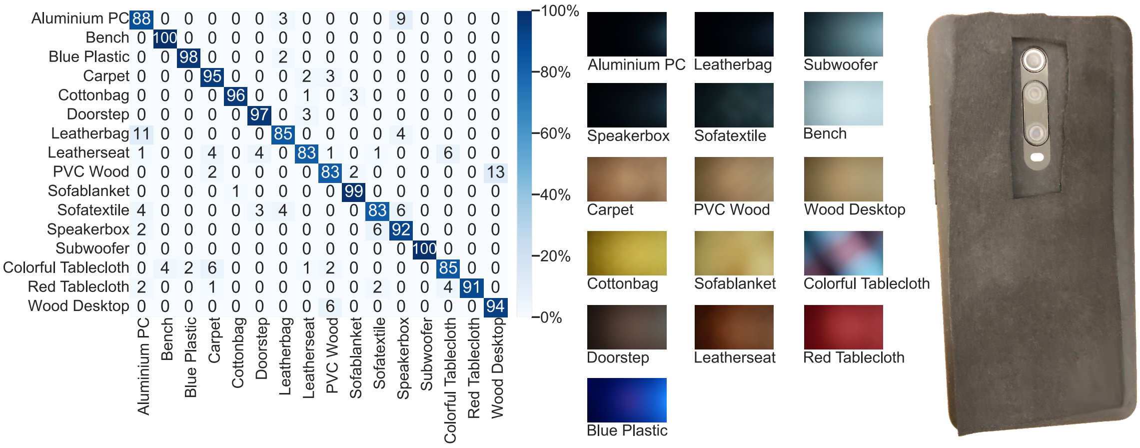

Figure 7.5: Confusion matrix with internal LEDs and several household materials taken on a Xiaomi 9T pro smartphone with a neoprene cover on the back to avoid interference of external light sources.

7.7.2 Notification Management

The smartphone is an ubiquitous communication tool. During office work, interruptions by phone notifications can cause an decreased productivity [Mes16]. However, one might use their smartphone for business and private tasks. Therefore, it may also be important for a user to respond to incoming notifications. However, it is not always obvious in which context the smartphone is currently used. Therefore, detecting the office desk could help to identify if a user is currently working. It would then be possible to control notifications and achieve a more focused work.

Figure 7.6 depicts the proposed idea. In A, the user manages private and business related contacts. The user can add new surfaces and related contexts in B. A surface is then captured in C by placing the smartphone on the surface and taking images. By taking several images and sending the images to a server, a SVM can be trained that recognizes the context. In D a user is sitting in front of the office desk where only business related messages are displayed on the smartphone. The user can see on the display who is calling or what message arrives. Private messages are not displayed in order to achieve a more productive working session. When the user places the smartphone on the office desk, the gyroscope can be used to recognize that the smartphone is not moving. Then an image can be automatically captured that is sent to a server where the surface is recognized. The classification result is sent back to the smartphone and used to manage incoming notifications. The same can be performed when the phone is placed on the coffee table during a break. In this case, only private messages are presented. As a prototype, the proposed technology has been implemented with the twitter API v2 to display different tweets from twitter accounts as shown in F [Twi].

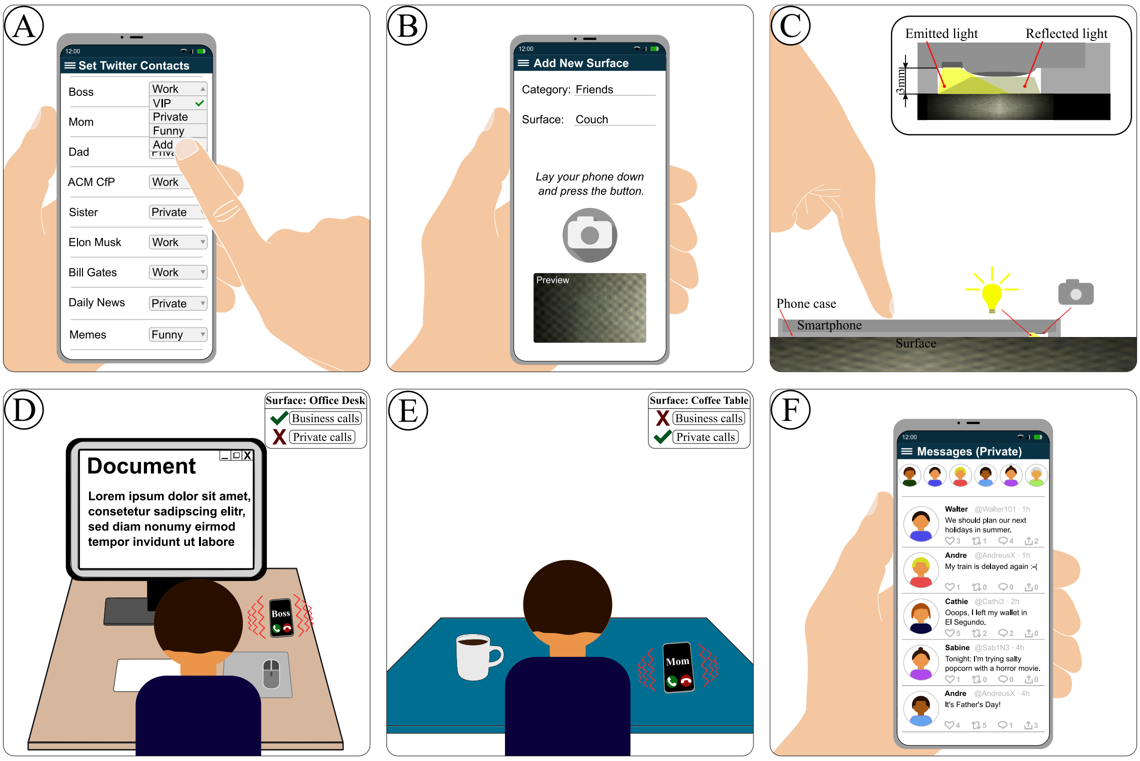

Figure 7.6: Story board for notification management with twitter: In A twitter accounts are linked to different contexts and surfaces. B shows the interface for adding new surfaces and C shows how the surface is captured. In D a user is working at the office desk while only business related messages are presented while in E only private notifications are shown during a break. In F the interface for displaying tweets is shown.

7.7.3 Facilitating Smartphone Search with Surface Information

Nomophobia is the fear of being unreachable for contacts without the cell phone [Yil15]. Misplacing one's smartphone may cause discomfort and can even lead to panic behavior for some users [Fur21]. The personal smartwatch can help to locate the smartphone when the user is within Bluetooth range. For instance on the Apple Watch, users can trigger ping tones to find the own smartphone. However, this feature is barely helpful in finding the phone in case of hearing loss. Therefore, a prototype was implemented that displays additional information of the smartphone's location on the smartwatch. By displaying the material on which the smartphone is lying on and a photo of the front camera on the smartwatch, a user can easier locate the personal phone in a home environment. Figure 7.6 depicts the proposed idea which has been implemented as a prototype. In A the user is looking for his smartphone in a home environment. He triggers the search function on his smartwatch in B and the phone starts ringing in C. In D, the triggered process is shown. The smartphone automatically takes an image of the front and rear camera and then classifies the surface material. The estimated surface together with the surface color and front image is sent to the smartwatch in E. By this, the person can easier locate the smartphone. When touching the device in F the phone stops ringing.

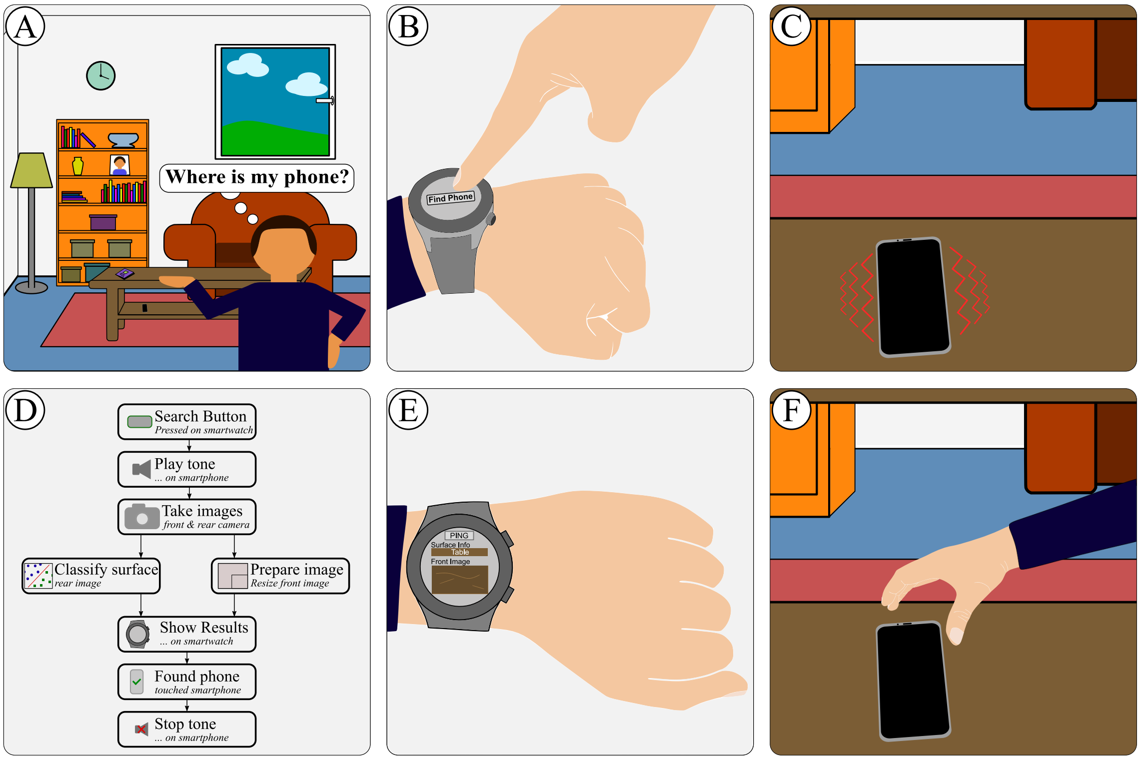

Figure 7.7: Story board for improved smartphone search. By displaying smartphone location information on the watch in E the user can easier find the smartphone.

7.8 Conclusion & Discussion

This chapter reported the use of built-in hardware of smartphones, namely the rear camera and white flashlight LEDs, for material sensing. Based on the measured RGB values at different intensity levels and LED combinations, high recognition rates of approximately 99 % were achieved on a dataset of 13500 images, consisting of 30 materials with 30 samples of each material and each sample consisting of 15 images taken at different illuminations. Due to API limitations in controlling the intensity of flashlight LEDs of current mobile phones [App18; Goo18], external LEDs, providing the same spectral distribution as built-in LEDs, were connected to a microcontroller and Bluetooth unit and placed in a housing.

The results of external flashlights show that only a few images per surface are sufficient to achieve acceptable recognition rates for typical surfaces. However, black surfaces turn out to be difficult to recognize with at least 89%. We had also various white surfaces in our dataset (*CardboardWhite*, *PlatePorcelain*, *ThileWhite*, *Paper*, *WallWhite*, *Whiteboard*), which proves that our proposed approach is not only based on color separation. The results in Table 7.1 indicate that built-in dual flashlights can reduce the number of images to be taken for material sensing. With two images (warm and cold white illumination) at 100 % intensity comparable recognition rates to SpeCam [Yeo17] were achieved. With our approach the time for the material sensing process can be reduced. Furthermore, we could half the number of calculated features from 768 [Yeo17] to 384. If the API restrictions were removed, the test could be repeated with the embedded flashlight and the application could be made available to a wide range of smartphone models. The case could then be replaced by a shielding around the camera and flashlight. The opportunity of keeping the display visible enables to show the sensing process as well as the results instantly. Feedback could be given when external light interferes with the sensing process similar to camera-based heartbeat measurements [Pel10].

Subsequently, internal flashlight LEDs were evaluated by capturing images of 16 household materials. For each material 30 images were collected resulting in a total of 480 images. We observed an reduced accuracy of about 92 % by using built-in flashlight LEDs. The higher confusions resulted from similar material colors and light reflections of several household materials. Thus, it can be concluded that heterogeneous color distributions are useful for deploying applications that use surface information.

Two examples were motivated and implemented in which surfaces can be integrated. At first, an approach for context-related notification management was presented. By using the twitter API, twitter accounts could be linked to the material where a smartphone is lying on. This interaction technique can serve to enhance productivity during work by suppressing messages and related notifications of private contacts. Furthermore, it would be possible to adjust the notification modes through different colors on an office desk. For this purpose, a user could position sheets of paper in different colors on the table. Depending on where the smartphone is placed, the user can specify his or her reachability and easily view the current state. Due to restrictions in fully managing all system notifications with the Android API [Goo18], we limited the example application to twitter messages. However, to increase the productivity of workers a full management of notifications is necessary. Besides the proposed system, it is currently possible to mute notifications at certain times in the Android system. This does not offer the opportunity of identifying the context in which a smartphone is being used.

Secondly, an approach was presented to enhance the search for a smartphone by displaying additional information of the phone location on a wrist-worn device. The classified surface, it's color and a front camera image was sent to the smartwatch to give the user hints on the phone's location. The proposed method aims to support the search of a phone in cases of hearing loss or muted phones where played tones do not support to locate a phone. Bluetooth 5.1 added the functionality of direction finding between two connected devices [Blu]. This feature hasn't been implemented in the proposed test application but could further support such searching tasks. By showing the relative direction of the watch to the phone by displaying with a small arrow on the watch display the search space could be further narrowed down. Even without a smartwatch, users could find their smartphone more efficiently by location-based features. For this purpose, the user could trigger a search event by a conversational interface. Here the user speaks loud in a room *"Hey phone, where are you?"* and the smartphone replies *"I'm lying on a light brown wooden surface and see a ceiling lamp over me. I will now play a ringtone for you."*.

A limitation of the presented material sensing approach is that users must first add surfaces to a dataset. However, in an environment with a heterogeneous material color distribution this process can be simplified. Generally, we conclude that surfaces can contain context related information that are able support users in different tasks.

CHAPTER 8

Facilitating Book Search by Surface Features

In the previous chapter, we found that a heterogeneous color distribution is beneficial for distinguishing surfaces. A home or office environment can consist of many different colored surfaces that enable to locate the phone or to recognize the context of smartphone usage. One particular case where such a color distribution can be found are book covers in bookcases. Finding a book can be difficult if the location on the shelves is unknown and the books are not sorted. Such bookcases can be found not only in home and office environments, but also in urban areas. Public bookcases integrate central sharing spaces into local communities and are also attractive to online book sharing communities. Bookcrossing is a form of book sharing where a book is placed somewhere in the wild. Public bookcases increase the probability of successfully sharing a book via Bookcrossing. However, searching for a particular book in public bookcases can be challenging because it is might not known whether the book is still available and the shelves are unsorted. This chapter investigates whether book search can be facilitated by augmented reality. A smartphone application is presented that recognizes books and bookshelves in camera image and subsequently highlights a desired book in the view. With our app, we investigate the following research questions:

> Can book search be facilitated with augmented reality?
>
> Are color and text features on the book spines sufficient to find a book on the shelves?

This chapter is based on the master thesis by Thilo Schulz [Sch19b] and a full paper at *MobileHCI 2020*.

- Schrapel, Maximilian & Schulz, Thilo & Rohs, Michael: ‘Augmenting Public Bookcases to Support Book Sharing’. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services.* MobileHCI ’20. Oldenburg, Germany: Association for Computing Machinery, 2020

Figure 8.1: Example public bookcase: The modified old phone booth protects the books from rain and attracts pedestrians. The inventory is highly dynamic and includes different genres from novels to specialist literature.

8.1 Introduction

Many books are read once, or not at all, and then left unused on personal bookshelves for years. A more sustainable use is achieved by book-sharing economies, which make read books available in public places. People can take and deposit books as they please, free of charge, anonymously, and without any administrative overhead. Public bookcases are enclosed book shelves located in public space that enable the effective sharing of books as a resource and are one example of the emerging sharing economy. The inventory of public bookcases is highly dynamic, as it is unpredictable which books are added or removed at any time. This makes public bookcases particularly valuable in the serendipitous discovery of interesting books. On the other hand online book-sharing communities like Bookcrossing[1] enable searching for particular books placed at any location in the wild. The journey of released books can be tracked by

1 www.bookcrossing.com

registered users. However, public decentralized sharing is not always successful, as books can get lost or thrown away. For this reason, online registered books can often be found in public bookcases. Nevertheless, it is difficult to search for specific books or to look for books in one's areas of interest, respectively, on public bookcases.

To facilitate the discovery of books, especially for a centralized public sharing scenario, we developed a mobile augmented reality (AR) application. It identifies and highlights relevant books in the camera view. The app thus provides a way of connecting the physical with the virtual world. A core problem that the app solves is the recognition of books in the shelf, which are typically not sorted in any way and might be oriented in different ways.

In this chapter we review related work on sharing communities and book sharing in particular. We present an algorithm that reliably recognizes books in book shelves based on their spine and discuss how graphical output is overlaid in the camera view to highlight books of interest to the user. Then we report on a lab-based user study that compares unaided visual and AR-supported visual search in book shelves, which shows that AR-supported visual search outperforms unaided visual search. We contribute a dataset of a large number of images of books in nine different public bookcases. The images were taken in different weather and lighting conditions. The dataset was used in evaluating the book recognition performance of the presented algorithm. Furthermore we interviewed users of public bookcases in order to learn about their motivations for using them.

The contributions of this chapter include (1) technical support for freely sharing books that only relies on a mobile application and general information on book covers, but does not require additional infrastructure, which would introduce maintenance costs; (2) an algorithm that recognizes books in shelves in different orientations and lighting conditions, and a way for overlaying graphical output; (3) a dataset of book images from public bookcases; and (4) several findings about the motivations of users of public bookcases.

8.2 Related Work

8.2.1 Sharing Communities

The sharing economy grows. Not only digital resources and services but also physical objects are shared, as exemplified by shared vehicles [Bar12a] of different kinds. Sharing is a way of using resources more economically. It can contribute to a more sustainable society that consumes less and produces less waste. Sharing is present in the commercial space as well as in non-profit contexts. Individuals regard the act of sharing as a positive behavior that increases happiness and the feeling of connectedness with others [Lig14].

Dillahunt et al. [Dil17] provide a detailed survey of research on the sharing economy in computing in general and in HCI in particular. They outline un- and underexplored aspects of research in this area and suggest future research directions. Fedosov et al. [Fed18] analyze how digital sharing economy services have expanded the range of physical and digital resources that can conveniently be shared among individuals. They also discuss sharing practices, motivations to share, privacy and trust issues, and design implications of sharing economy services.

Light and Miskelly [Lig14] explore how sharing takes place in everyday life and how digital services support it. They report on “One Small Thing,” a public bookcase in a neighborhood in London. The bookcase is located on a busy street in a former BT phone box. Books are regularly taken out, replaced, and exchanged by local residents. Light and Miskelly give an account of the person who established the bookcase and his doubts regarding the prospects of such an unattended facility. No formal organization takes care of the public bookcase, but local people provide minimal maintenance. Bulk deposits replenish books being taken out over time. Apart from digital documentation of its existence on Facebook, the public bookcase is not supported by digital technology. The experience has been that the public bookcase has survived in good shape and that the free sharing model incurs goodwill that protects it.

8.2.2 Public Bookcases

There have been two main ways of sharing private books publicly. First, books have been made available at a designated location with protection from the elements. Secondly, books have been left behind at arbitrary public places, like a park bench, typically protected in a labeled plastic bag. Both approaches differ from traditional libraries in that everyone can donate books as well as take books and that there is almost no administrative overhead involved. There is also no clear ownership of the donated books. Readers can keep a book as long as they want.

“BookCrossing” [Dal14], launched in 2001, is an initiative that mainly follows the second, decentralized approach: Books may be deposited at anywhere for others to pick them up or they may be handed over personally to another individual. The aim is to “make the whole world a library”. After registration at the BookCrossing[1] site users can enter information about books they would like to share into an online database, create a label with a unique “BookCrossing ID”, and attach it to the book cover. The ID can be used to track the journey of a book via registered readers. Although this type of sharing is not always successful, as books can be placed anywhere in the public and might get lost or trashed, in 2020 there were already 2 million registered users who shared 13 million books.[1] Nowadays, many books from the BookCrossing community are also placed in public bookcases, as this increases the chance to successfully share a book.

1 www.bookcrossing.com

Public bookcases represent the more centralized approach to publicly share free books. First “free open-air libraries” were established in Germany and Austria at the beginning of the 1990s, initially as artistic acts [Han16]. Books are placed at a fixed location and no database is kept of the inventory. The advantage is that anonymous public sharing is possible, at the disadvantage that volunteer mentors have to permanently check the stock and quality of the books as well as the bookcase itself. Now public bookcases can be found all over Europe. “OpenBookCase”[1] provides a worldwide map of public bookcases. The worldwide success of public book sharing started in 2009 when Todd Bol built a small bookcase and placed it in his front yard in Hudson, Wisconsin. In 2012 he founded “Little Free Library”[2] (LFL), which is a nonprofit organization that promotes neighborhood book exchange.

There is a small research community that investigates the phenomenon of public bookcases. Rebori and Burge [Reb17] suggest using geospatial analysis to highlight potential locations for new LFLs. However, since studies about the usage are often carried out locally on a small scale and are not standardized, no representative usage figures can be given. According to estimates of 20 volunteer mentors from a large German city, on average at each public bookcase about 75 to 121 books are taken out per week and mentors dispose about 11 to 18 defective books per week [Cla16]. In a questionnaire with 66 users of different age groups at three bookcase locations, it was found that typically more books were inserted than withdrawn. The surveys have also shown that sharing via public bookcases is a good source of knowledge, especially for socially disadvantaged groups of the population. From the collected data it was estimated that about 2.8 % of all book purchases can be avoided [Cla16]. Generally we note that in the area of public bookcases there is still much potential for future research [Sno15].

8.2.3 Visual Search

In the past decades *visual search* has been investigated in detail, as it is an important everyday task [Cha13]. Unaided[3] visual search for books in bookcases is a perceptual and cognitive task that requires distinguishing visual features of the *target* book from visual features of other *(distractor)* books [Tre80]. We use the term *mobile visual search* to refer to visual search that is supported by camera-equipped mobile devices with AR functionality. AR applications for mobile visual search allow the user to perform search queries about objects in the camera view [Gir11]. Mobile visual search does not rely on instrumenting objects, e.g. with QR codes, but rather uses visual features of the objects themselves for identification. The advantage is that no instrumentation of the objects is necessary, but the recognition task gets

1 openbookcase.org
2 littlefreelibrary.org
3 We use the terms unaided and manual visual search interchangeably.

more challenging.

The visual features of the objects can be extracted and classified either directly on the mobile device or on a remote server. Applications such as *Bing Visual Search* [Mic19] or *Google Lens* [Goo19] use online deep learning methods to identify objects or translate texts. The recognition results are used as query arguments for finding visually similar items. However, these services do not allow searching for a particular book among other books on a shelf. Tsai et al. [Tsa10] extract CHoG features from DVD and CD covers that are compared with an online database to provide the user with additional information about the item.

Mobile visual search in bookcases has been explored before in the context of large libraries. Yang et al. [Yan17] present an online approach for reducing time and labor involved in inventory management. By using the Hough transform for extracting book spines with subsequent text recognition, based on a deep convolutional and recurrent neural network, they identify books on shelves. Matsushita et al. [Mat11] use projectors and a camera-based tracking system above the shelf to highlight individual books and to display further information. They found that search times for a sorted shelf with Japanese books in different colors were not affected by displaying additional information. Similarly, Löchtefeld et al. [Löc10] use a portable projector and a mobile phone camera to highlight individual products and books in a shopping and library scenario. Book ratings were displayed over the object in a dark environment due to low brightness of the projection. Scenarios and their options for feedback were analyzed, but the search performance was not examined. Lee et al. [Lee08] developed a system for libraries to sort books on a shelf. Book spines were identified by color features in a heterogeneous color distribution at an accuracy of 96 %. Quoc and Choi [Quo09] developed a framework for segmenting books and characters. By applying high frequency filtering and Canny edge detection they segmented books and characters at an accuracy of 93.3 %. To display metadata, such as book price and rating to the user, Chen et al. [Che10] developed an AR application that analyzes parts of a bookcase. While their system offers an acceptable latency of 1 s, even with low computing power, only a small part of a bookcase can be analyzed at once and only if all of the books have the same vertical orientation. Nevetha and Baskar [Nev15] use a line segment detector [Gio12] to also detect slightly inclined books up to an angle of 15°. In a test shelf of 20 books, 11 books could be successfully identified. In order to solve the problem of orientation and view perspective, Talker and Moses [Tal14] developed a segmentation algorithm that achieves an accuracy of 82 % which was used to reorganize books on the shelves sorted by height, width, or color.

In contrast to related work, public bookcases like shown in Figure 8.1 present a major challenge. Any orientation, size, color, and even language can occur. The structure of the bookcases can also be widely different. Books can be damaged or have reflective covers so that text recognition is difficult. Perspective distortions can occur when using a handheld camera phone. Moreover, any costs, such as for the use of servers or their maintenance, should be avoided as typically no funding is available for public bookcases and digital services around them. Consequently, the mobile app must operate independently of a server infrastructure and allow for text or voice input to specify desired books or book types. In addition, the system has to find a book faster than unaided visual search.

8.3 Book Recognition

Figure 8.2 gives an overview of the steps involved in recognizing book spines. First the title of the book is entered via the touchscreen keyboard or via voice input with Android's speech recognition API [Goo18]. Then a high-resolution image has to be taken that covers the book spines in the bookcase. This is done when the phone's motion sensors detect slow movement (to prevent motion blur) and either the user explicitly presses the shutter button on the touchscreen or a pretrained *MobileNet* [How17] detects the bookcase in the video stream. The video stream is accessed via Android's *Camera 2* API [Goo18]. The MobileNet is run on the phone via *TensorFlow Lite* [al15]. Next, line segments are extracted with *OpenCV* [Its15] functions, which are used to detect shelves (i.e., individual compartments in the bookcase) and orientations of book spines. The color information of the recognized books is then used to preselect those that are similar to the target. Subsequently, the title, subtitle, author, publisher, and color information of each remaining book spine is extracted and compared to the desired book, of which all features are stored in a database on the phone. For text recognition *ML Kit* [Inc19] is applied. Finally, the position of the target book is highlighted with a green frame and a short 300 ms vibratory feedback is given. *ARCore* [LLC19] and *Sceneform* [Inc18] are used to augment the video stream. In the following, we give a detailed description of our algorithm and name aspects that have to be considered for public bookcases.

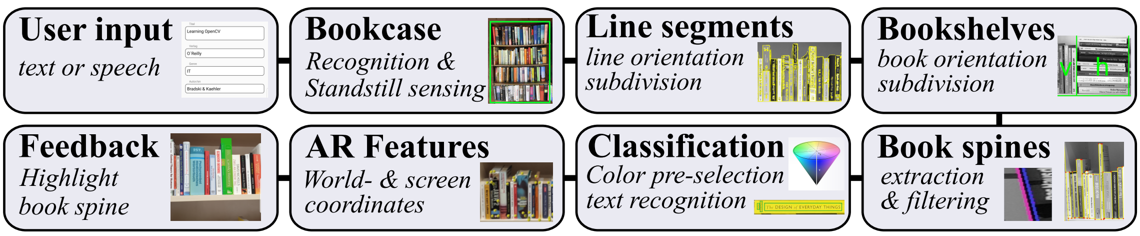

Figure 8.2: Book spine classification algorithm: The user specifies the book title via text or voice input. From the camera image the algorithm extracts line segments in various orientations, which are used to detect outlines of shelves and spines. Text and color information on the book spines is used to recognize books. Each detected book is highlighted in the camera image with a semi-transparent frame.

8.3.1 User Input

The user initially has to specify the desired book or books. This can be done using any fragment of publicly available bibliographic meta information. In our prototype the user enters the title of the target book via on-screen keyboard or voice. Although voice input offers a simple and convenient mobile text input method [Rua18], voice input may not be appropriate in public space as it can be perceived as disturbing. Moreover, in noisy environments like pedestrian walkways, speech recognition can be erroneous and cause mode errors [Rua18], therefore traditional mobile text input remains equally important.

In order to correct input errors, we use the normalized Levenshtein distance [Lev66] which specifies the minimum number of delete, insert, or swap operations required to convert one string to another in the mapped interval between 0 (no similarity) to 1 (equal). The resulting list of book titles is sorted by decreasing similarity. Due to expected speech recognition errors, especially on foreign language titles, all book titles with a similarity of more than 0.4 are added to the list, if no equal title was found. For text input, only titles that match the typed characters are displayed. If no appropriate title can be found, the same method is used as for voice input.

8.3.2 Bookcase Recognition

Once the desired book has been selected, the visual search can be started, either explicitly by pressing a camera button or automatically when the bookshelf is recognized. We use the model *Mobilenet_V1_1.0_224* [How17] which is pretrained on 1000 classes, including bookcases. The low average latency (160 ms), high recognition rates (top-5 accuracy: 89.9 %) and relatively little memory (16 MB) requirement are advantageous on the mobile device. Since motion blur would lead to illegible book titles, camera images are only processed when the camera is kept still. Our prototype uses the phone's gyroscope to detect motion: If the sum of squares of the three axes is below a threshold then an image is captured on which the search algorithm is subsequently executed. An appropriate threshold for the tested *Samsung Galaxy S9* is $T_{Gyr} = 0.001$.

8.3.3 Book Spine Segmentation

For the book spine segmentation, we first extract line segments, which are then used to detect book shelves and book spines in any orientation.

8.3.3.1 Line Segment Detection

To recognize line segments in the image, the parameter-free bottom-up line segment detector by Gioi et al. [Gio12] is a suitable choice. Unlike other approaches [SRa15] this algorithm implements a fast and precise segmentation process for the rather small line segments of books in the large camera image. To decrease noise sensitivity and to eliminate line segments due to text, a threshold is used that discards short lines of 25 pixels or shorter. Subsequently, all lines are categorized as either horizontal or vertical.

8.3.3.2 Shelf Detection

The line segments are used to subdivide the bookcase into individual shelves. When a horizontal line is longer than 50 % of the image width, it is categorized as a shelf boundary. This approach achieves good results for images in which the bookcase almost fills the image. At larger distances the text size on the spines becomes too small to be recognized. Subsequently, the Cohen-Sutherland algorithm [New79] is used to create rectangular frames around the shelves with book spines and associated line segments.

8.3.3.3 Book Spine Orientation

Especially in public bookcases, book spines can appear in any orientation. The minimal assumption regarding book spines is that their heights and widths differ strongly. Based on this assumption, a rough subdivision into horizontal and vertical sections is first carried out on the bookshelf by analyzing the line segments. For all vertical lines, the average length is calculated, as well as for all horizontal lines longer than 20 pixels. Shorter lines are often caused by text or graphics on the spine. Since the distance between two horizontal lines is smaller for standing books, vertical and horizontal regions can be extracted using a threshold. As a limitation of this approach, empty areas are also classified as vertical or horizontal book regions, but because of the subsequent text recognition, these do not result in a match. Finally the vertical and horizontal regions are clipped with the Cohen-Sutherland algorithm [New79].

8.3.3.4 Bounding Boxes

From the previously segmented areas with vertical (based on line segments at angles between 45°and 135°) and horizontal books, each individual book spine can now be extracted. Mixed orientations, e.g., standing books on a lying book, are recognized as pure vertical areas in which the horizontal book is not recognized. For simplification, horizontal regions are rotated by 90°so that each book spine has a vertical orientation. If all of the books are aligned exactly within the areas without any perspective distortions, they can be segmented by combining vertical and horizontal lines to bounding boxes.

Small angular differences of less than 1°of text between two adjacent lines are connected, because larger differences between adjacent segments indicate other features of the book, such as text, graphics, or damage and wrinkles, which are discarded. Perspective distortions as shown in Figure 8.3 must also be eliminated when looking at the entire bookcase, which may contain edges along the top of the books. A line is therefore rejected if its lowest point is at a higher position than the

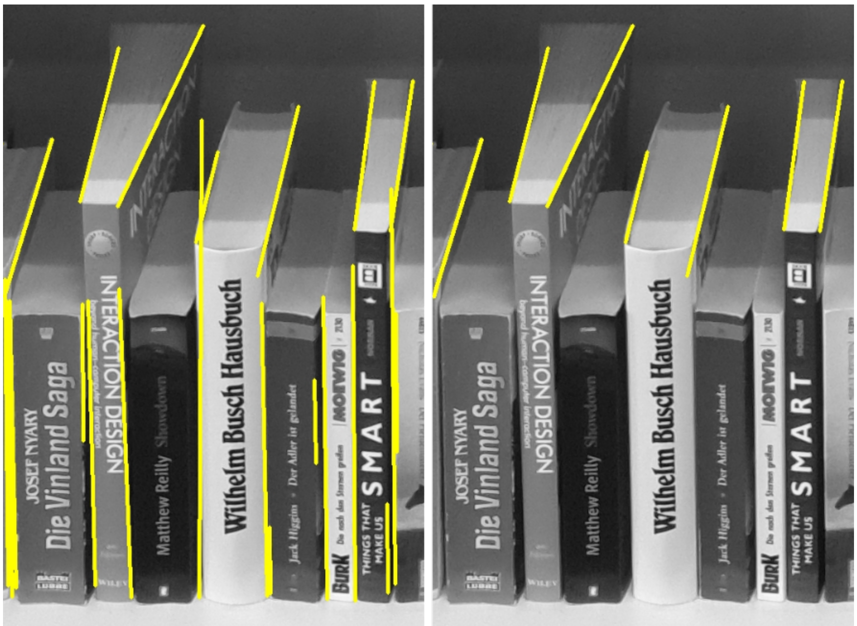

Figure 8.3: Line rejection: Initially extracted lines (left) and lines to be rejected (right). The lower endpoints of the rejected vertical lines are subsequently used to create the bounding boxes.

median of all vertical lines.

Finally, bounding boxes are created at the intersections of the horizontal line segments and the extracted vertical lines. While lower boundary lines in vertically arranged books are extracted by connecting the endpoints, the different lengths of the books in upper horizontal lines or lying books require a more sophisticated analysis of the segments. The surrounding horizontal line segments are followed to the end points of the vertical boundaries, which best connect the end points.

8.3.4 Feature Extraction

To find the searched book it is necessary to extract visual features of the spines. In general, books are analyzed in two steps. First, books are rejected whose color is too far from the target item. Secondly, after applying text recognition in all four orientations, books are rejected if the corresponding Levenshtein distance is too large.

8.3.4.1 Color Information

To extract the color information we use the HSV color space. The *hue* circle is coarsely divided into color intervals to make the color information more resilient to varying lighting condition: Red [320°, 40°], yellow [40°, 80°], green [80°, 160°], cyan [160°, 200°], blue [200°, 280°], and magenta [280°, 320°]. A pixel is determined as black if its brightness *value* is less than 0.15 % and categorized as white/grey when the *saturation* is below 0.15 %. For normalization, all $n = 8$ feature intervals are subsequently divided by the number of pixels of the book spine.

To determine the color similarity s, the respective features b are compared with the ones of the target book spine a. By squaring the resulting distances, negative values are avoided and large differences are weighted more strongly. To obtain a normalized similarity value s in the interval $[0, 1]$, the accumulated distances of all features are subtracted in Equation 8.1.

$$s(a,b) = \max\left(0, 1 - \sum_{i=0}^{n-1}(a_i - b_i)^2\right) \tag{8.1}$$

Since the books in public book shelves can possibly be in a poor condition, only books with a similarity greater than 0.5 are considered for subsequent text recognition.

8.3.4.2 Text Recognition

Text can appear in different orientations on the spine of a book. Therefore it is necessary to extract the text in all four directions. We use for this step *ML Kit* [Inc19] as it performs well and is freely available. Another challenge occurs when the author, publisher, and book title are printed in the same orientation. Then multiple strings are may merged together. For this reason, the extracted text is compared to the target in substrings, again using the Levenshtein distance. The target information consists of publisher, author, title, and subtitle. The maximum resulting similarity is subsequently stored for each target information and added to an overall similarity as in Equation 8.1.

8.3.5 Classification

In order to finally determine the position of the book on the shelf, the five similarity values, consisting of title, subtitle, author, publisher, and color information, are summarized and weighted. The weighting is necessary as, e.g., not every subtitle is noted on the spine.

To estimate adequate weights, a bookcase (5 shelves, 1 m × 2 m) in a laboratory environment was filled with 201 books from public bookcases in different orientations. 50 books were randomly selected and search queries initiated on a *Samsung Galaxy S9* at a fixed position where the camera image completely covers the bookshelf. In each of the 50 trials 201 book spines were recognized and their corresponding color and text features stored in a database. Then a random forest classifier was trained for 1000 times on this dataset consisting of 50 positive and 10000 negative samples, with a ratio of 20% randomly picked test data. The negatives samples represent the 200 extracted book spine feature vectors from each search query except the target book. The estimated feature weights of all 1000 random forest classifiers were then averaged. The resulting weights are used to calculate the overall similarity of a book spine to a target in Equation 8.2.

$$S = w^T s = \begin{pmatrix} 0.6436 \\ 0.2233 \\ 0.0789 \\ 0.0295 \\ 0.0247 \end{pmatrix}^T \begin{pmatrix} s_{\text{title}} \\ s_{\text{author}} \\ s_{\text{color}} \\ s_{\text{publisher}} \\ s_{\text{subtitle}} \end{pmatrix} \tag{8.2}$$

The calculated scalar product S weighs the title most strongly, followed by the author. Textual information, except for the title, is not always noted on the spine of a book or may be too small to be captured well in the camera image. In addition, the average accuracy on the test data was over 99.94 %, which shows that in a laboratory environment the feature vector describing the book spine is distinctive enough to distinguish a target book spine from other books. However, since the test was performed on books from public bookcases, it is to be expected that features, such as author and title, greatly differ between the book spines. In a library scenario with books sorted by author, the above weights would may not be applicable. In a public bookcase the highest resulting scalar product S of all analyzed books is chosen as the desired book and highlighted in the last step.

8.4 Visual Highlighting

To show the result of a successful search to the user, the position of the book spine is highlighted in the video stream. *ARCore* [LLC19] and *Sceneform* [Inc18] are used to keep the bounding box locked to the same book spine.

For this purpose *ARCore* uses concurrent odometry and mapping [Esh17], which combines feature points from the camera image with inertial measurements. During this procedure the algorithm estimates trackable planes in the image so that afterwards virtual objects can be attached to anchors. *Sceneform* then allows obtaining screen coordinates from world coordinates derived from an anchor in order to establish a connection between the two. During this step, position features are extracted from the camera image and at the same time the book search algorithm is executed. To display the position of the book as accurately as possible, at least 20 position features should be available, which is typically the case due to the heterogeneous properties of book shelves. After classifying the desired book spine, the closest anchor is moved into its center in order to highlight the position. In *Sceneform* a transparent green rectangle is drawn on the spine of the book which is 50 % wider than the spine itself to make the highlighting easily to see even for slim books. The subsequent movement of the camera towards the selected object leads to small deviations in the highlighting position, as can be observed in Figure 8.4.

Figure 8.4: Debug view of a highlighted book with displayed AR feature points: Angular differences are also taken into account. There are minor deviations from the center point caused by the movement of the camera and the resulting deviation of the anchor points.

8.5 Dataset

While public datasets, such as by Iawa et al. [Iwa16], are limited to book covers, to the best of our knowledge no datasets containing book spines are freely available. Moreover, the condition of the inventory strongly varies in public bookcases, which possibly results in less accurate text recognition. A meaningful dataset of 767 randomly picked books from public bookcases were included in our dataset with different sizes, colors, reflective properties, languages (English, French, German, Polish), topics, and conditions. The derived features are stored on the phone with a total size of 160 kB.

The complete collection consists of 317 books, which we used in a lab study, and 450 books from an evaluation in the field. A csv-file provides bibliographic meta data, such as genre, rating. Taken together, the books fill more than four bookcases (1 m $\times$ 2 m) of four shelves each. After our studies, all books that we took from public bookcases were released back into the public bookcases. The entire dataset consists of 767 book spines including novels (66 %), thrillers (10 %), children's books (8 %), education books (4 %), politics (2.8 %), advisors (2 %), history (1.9 %), tourist guides (1.6 %), cooking (1.3 %), poems (1 %), religion (0.4 %) and others (1 %). In addition, images of the public bookcases and closeups of the individual shelves serve to evaluate and compare algorithms for future research. It is available online together with our Android application at: `https://github.com/M-Schrapel/Public-AR-Booksearch`.

8.6 Experiments

The goal of the lab user study and the technical field evaluation is twofold. We wanted to compare the time required to locate a book using the mobile AR app (mobile visual search) versus unaided visual search. We also wanted to estimate the accuracy of the proposed algorithm on search queries performed on realistic images. In addition, we intended to clarify whether the application offers a perceived benefit for the target users and understand further factors regarding the use of public bookcases.

8.6.1 Lab User Study

The aim of the first study was to evaluate whether the proposed algorithm is faster than the manual visual search. To focus on the core aspects of the algorithm and to exclude random factors, the study was conducted in the lab. The study took place in a quiet room with 15 volunteers (12 male, 3 female) aged 20 to 65 (mean age $\bar{x} = 30.4, \sigma = 13.7$).

We used two identical adjacent bookcases. Each one had a width, height and depth of $w \times h \times d = 100 \times 200 \times 40$ cm (see Figure 8.5) and each had five shelves, located at heights 10, 48, 87, 130, and 170 cm from the floor. Since none of the public bookcases that we found had books at the bottom shelf, probably for ergonomic reasons and to prevent soiling, we left the lower shelf empty. In each of the two bookcases the second shelf from the top was filled with a series of 44 books sorted by title, to evaluate the search performance for sorted shelves. For the first part of the study the volunteers were not informed, and not aware, that the books in one of the shelves of each bookcase were sorted.

To compare digital and manual search, one bookcase was covered as shown in Figure 8.5 and the study either began with the digital or the manual search method. Both the initial bookcase and the initial method were counterbalanced. For the digital search, the mobile AR app was first explained to the participants through several practice trials, which were performed on a printed image of another bookcase with a different set of books. Each participant was instructed to touch the desired book to complete the search. One trial of the experiment consisted of the experimenter announcing the randomly picked title of the target book and measuring the time until the book was touched. The measurement was done with a stopwatch. The participants always started at a predefined position marked on the floor at a distance of 2 m from the bookcase. For the digital search trials the participants had to start the application on the phone and enter the title of the desired book using the on-screen keyboard. The start of the application was included as this is likely also needed in a real book search.

After 20 trials, the initial bookcase was covered and the search was repeated with the other bookcase using the other method (either digital or manual search). To evaluate the performance of voice input, the study was repeated with just 10 further trials on the bookcase on which the digital search with on-screen text input had been performed before, without additional manual search to minimize fatigue.

In the last part of the study, the participants were informed that the second shelf from the top in each bookcase is sorted by title. All participants said that up to that point they had not noticed that the books in these two shelves were ordered by title. Digital search with on-screen text input and manual search were performed for ten trials on these two sorted shelves in the same way as before. To obtain qualitative results, we asked for the preferred search method in a final questionnaire. The whole study took about 45 minutes on average.

Figure 8.5: Study setup: For each part of the study one bookcase is covered to minimize learning effects. The random distribution of the books' spine colors and sizes is similar as in public bookcases. The second shelf from the top is alphabetically sorted by title, which no participant noticed. The lower shelf was left empty for being located too low.

8.6.2 Technical Field Evaluation

The second study was intended to measure the algorithm's accuracy in the field and whether the target users potentially benefit from a mobile AR app. A further aim was to gain a better understanding of relevant factors surrounding the use of public bookcases.

From nine different bookcases in four cities and one village, 50 books of each bookcase were added to our dataset, resulting in a total of 450 books. The bookcases were located near pedestrian walkways, residential areas, city centers, in a super market (indoors), in a train station (indoors), and in a bus stop (roofed). The randomly picked book spines were photographed and added to the dataset. All photos were taken with an *Apple iPhone 8 Plus*. After five days the experimenter returned to each of the nine locations and every book of the first visit that was still present, was searched three times using our app prototype running on a *Samsung Galaxy S9*. The main goal was to estimate the recognition accuracy. In addition, the experimenter measured the illuminance (in lux) with a light meter and counted the number of removed books.

The tests were paused when a person intended to use the bookcase. During this interruption, the bookcase users were briefly interviewed about their personal use of public bookcases. They were asked what kinds of books they are looking for, which ones they put back on the shelves, whether they had ever searched for a specific book, if they take extra detours to get to the bookcase and how many books they take out per week.

8.7 Results

8.7.1 Results of the Lab Study

The mean search times for unsorted[1] vs. sorted books, manual vs. digital search, and text vs. voice input are given in Table 8.1. For the unsorted case, the time for digital search at 38.4 s is only about half (49.9 %) of the time for manual search at 76.9 s. At 36.9 s digital search with text input via soft keyboard is slightly (7.3 %) faster than text input via voice at 39.8 s. For the sorted case, manual search is much faster than digital search (7.6 s vs. 34.3 s), which confirms that the participants most likely did not realize that one of the shelves in each bookcase was sorted, before we revealed this fact to them. Again, for the sorted case manual text input was slightly (8.1 %) faster than voice input. The box plot in Figure 8.6, left, shows the strong differences between manual and digital search in both parts of the experiment.

Since the times for the search of a book are not normally distributed, a *Wilcoxon Signed-Rank* test was performed. The results show that the digital search with

1 "Unsorted" refers to the period before a participant was informed that one of the shelves in the bookcase was indeed sorted.

Table 8.1: Mean search times by condition: On unsorted shelves, the presented app achieves significant performance increases.

Method	**Unsorted**		**Sorted**	
	$\overline{x}$	σ	$\overline{x}$	σ
manual	76.9	51.4	7.6	5.7
application	38.4	14.5	34.3	12.0
text input	36.9	13.9	32.9	10.8
voice input	39.8	15.1	35.8	

text or voice is significantly faster than manual search in the unsorted bookcase ($z = -3.408, p = 0.0007 < 0.001$). There is no significant difference between text and voice input ($z = -1.25, p = 0.2115 > 0.05$), while in sorted shelves the manual visual search is significantly faster than digital search ($z = -3.127, p = 0.0005 < 0.001$). The sorted and unsorted cases are not directly comparable, because the sorted case only comprises a single shelf. A clear learning effect can be observed for manual search, as shown in Figure 8.6, right. The learning effect is expected, because as participants look for books they probably remember the locations of titles they have already seen in earlier trials. The participants achieved a total accuracy of 84.9 % averaged over both bookcases. The complete dataset was tested on a *Samsung Galaxy*

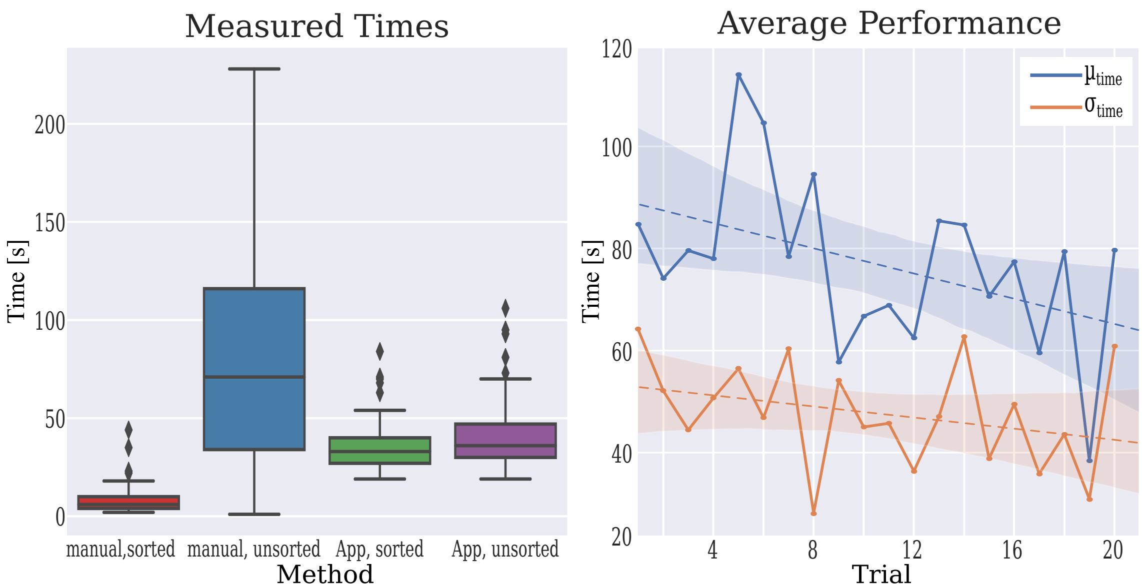

Figure 8.6: Times for visual search: Comparison of the search times for manual and digital search in sorted and unsorted shelves. Manual search in the unsorted shelf is substantially slower and more variable than the other conditions. Right: Mean time over trial for manual search. A learning effect is apparent. The σ-curve illustrates that the participants not only get faster, but also more accurate over the time.

S9 with a mean illuminance of $\bar{x} = 498$ lux ($\sigma = 80.1$ lux), measured at the shelves.

The times for the individual steps of the recognition algorithm are listed in Table 8.2. It becomes apparent that at 13.1 s the text recognition step took the majority of the time. One reason for this is that the text recognition is performed four times, once for each direction, for each candidate book spine in the image. This time can be reduced by making the color constraint stricter, but then false negatives (omissions of actual book spines) would become more likely.

Table 8.2: Mean times of each step of the algorithm: Text recognition requires the greatest effort.

Step	**Time [s]**	**Proportion [%]**
Line extraction	1.58	9.7
Bookshelf detection	0.02	0.1
Book spine orientation	0.01	0.1
Bounding boxes	0.46	2.8
Color feature extraction	1.12	6.9
Text recognition	13.11	80.4
Total	16.30	

Questionnaire responses of the participants are shown in Figure 8.7. In general, digital search is preferred over manual search. Manual search was experienced as more exhausting than digital search. Text input is preferred over voice input and perceived to be faster on unsorted shelves. The perceived performance is consistent with the measured performance.

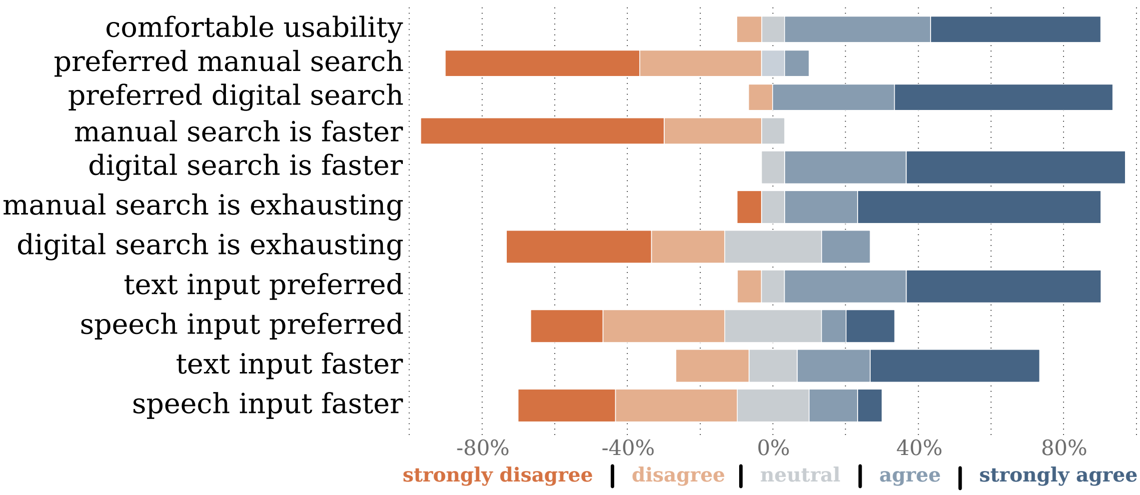

Figure 8.7: Qualitative results of the study: The participants prefer digital over manual search and text over voice input most. In unsorted bookcases, digital search was perceived to be faster than manual search.

8.7.2 Results of the Field Evaluation

In the field study, an average accuracy of 80.0 %($\sigma = 1.8$ %) was achieved across the nine public bookshelves. In within five days an total average of 59.2 %($\sigma = 14.4$ %) of the 450 books were taken out. Thus the application has been tested on 275 books, each of which was searched three times, which yields 825 trials. The light conditions varied strongly, from 380 lux to 13220 lux with a median of 4523 lux ($\bar{x} = 4295$ lux; $\sigma = 4093$ lux).

From the interviews with 16 female and 9 male pedestrians (mean age 48, $\sigma = 17$ years) it was discovered that 16 persons were looking for novels, 10 for thrillers, 6 for children's books, 4 for educational books including encyclopedias, 3 for advice books, and 5 for other topics, such as cooking or politics. On average each user visits the shelves once a week and takes additional detours of three minutes for taking out about one book on average. Three persons stated that they had never put a book on the shelves, while the others shared about 25 books in total. Eight persons stated that they had serendipitously discovered books of interest they were looking for, and one woman mentioned that she had just bought the same book online before. Four persons said that they were looking for their favorite authors, while one man remarked that there is a low chance finding a particular piece of literature. The other interviewees were generally more interested in discovering and reading unknown books.

As the interviewees only cover a very small part of the user population a Web crawler for Bookcrossing was developed to extract information about user demographics and shared books within the surveyed areas. The 868 registered users in the range of the examined bookcases had a mean age of 44 years ($\sigma = 13$ years) and shared 3900 books in total. In the book release entries, noted locations, and comments, we analyzed for keywords, such as "public bookcase," "compartment," and "shelf" to estimate the prevalence of centralized and distributed types of sharing. In 1400 collected cases, or 36 % of all gathered information, public bookcases are involved in book sharing. Thus, it can be assumed that the targeted users would benefit of such an application. We also looked at several threads in the Bookcrossing forum regarding public bookcases, Little Free Libraries and mobile applications. We could identify that users often ask for apps that address book sharing mostly related to finding and releasing books as well as book recommendations and scanning of Bookcrossing IDs. At this point when books are shared in public bookcases, an AR application can support the users. The Bookcrossing users stated that they are interested in knowing if a book is still at the place where they released it or are looking for it.

8.8 Discussion

8.8.1 Lab Study

The first study has shown that mobile visual search outperforms unaided visual search on unsorted bookcases. Limitations occur due to varying lighting and book conditions, small text, and segmentation errors. To find out which recognition errors occur, books in a shelf were placed in different orientations and photographed from different angles. Figure 8.8 shows typical examples of factors that deteriorate recognition performance.

In Figure 8.8a, the spine is extracted incorrectly due to perspective distortions and color similarities. In this special case the shadow coincides with the edge of the horizontal book and both books have similar colors. As a result the two books are not separated. In Figure 8.8b perspective distortion leads to the extraction of the top the book as a part of the spine. In Figure 8.8c shadows cause the spine of the book to be segmented incorrectly. As before, this leads to errors in color selection. Figure 8.8d shows errors caused by books in poor condition. The large book is not extracted correctly due to the torn cover and the graphics in the upper part of the spine. The same occurs with the spine at the right corner.

In Figure 8.8e the book spine segmentation is prevented by the protruding adjacent book. Secondly, the font is too small for a subsequent text recognition. In the case of the protruding books, parts of the book cover are recognized as additional book spines. Lastly, Figure 8.8f shows an example of two similar books with reflective covers. Similar titles led to misidentifications in the study. Another challenge is posed by the left spine. Due to light reflections parts of the book are not recognized. This shows that reflective books are challenging under certain lighting conditions, which also had an impact on the accuracies in the field study. The left book spine exemplifies mixed title and author information, resulting in lower text similarities.

These challenges show that book spines cannot be recognized flawlessly as environmental influences strongly affect performance. Features such as HoG [Dal05] are not applicable for public bookcases, because books can occur in various quality conditions and under varying lighting conditions the computed features would differ widely. For example, wrinkles can cause the detection of edges that are not present on the original book spine, which then significantly affects the classification. Therefore, it is important to only rely on robust features that also apply to books in relatively poor condition.

The questionnaire results show that participants definitely liked the idea of augmenting public bookcases and supporting visual search. Text input via soft keyboard was preferred, because voice input led to errors, especially when book titles in foreign languages had to be entered. For this reason, digital search via speech input was perceived as exhausting by a few participants, when they had to enter a book title twice. Text input was perceived to be faster. However, voice input is an accepted

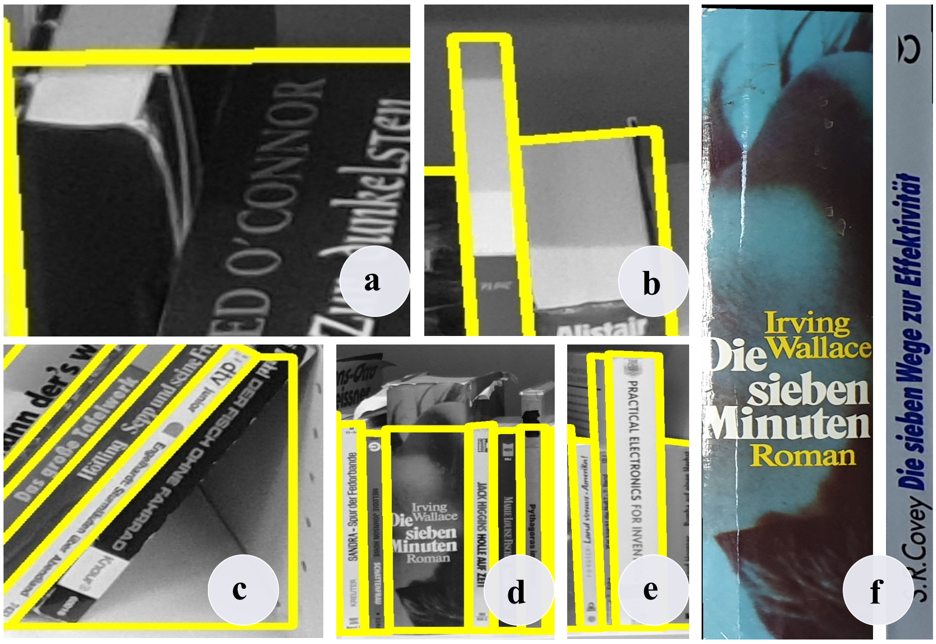

Figure 8.8: Book spine segmentation errors: (a) dark book spines and shadows at the same height, (b) incorrect segmentation due to perspective distortion, (c) incorrect segmentation due to shadows, (d) bad book conditions, (e) blurred text and protruding books, (f) similar titles and titles mixed with author and reflective book spines.

method on current smartphones, hence most study participants said that they could imagine using speech recognition if it were more reliable.

8.8.2 Field Evaluation

The effort for maintenance of server systems and databases should be kept low as very little or no resources are available to support public sharing of free books. In the wild, the maintenance costs of each bookcase are estimated to be only about 60 € per year [Han16], because mentors perform this task on a voluntary basis. During our field evaluation we observed that not only mentors, also other users feel responsible for public bookcases. When people brought books, they mostly placed them in an orderly vertical position on the shelf to ensure that no other books are hidden. This is beneficial for the presented mobile app, as this simplifies the recognition process. However, not every public bookshelf may have the same level of support as the observed ones.

Furthermore, it should be clarified for the desired use case how the presented application can be used together with online sharing, i.e. Bookcrossing. While releasing a book in the public bookcase the spine could be uploaded to Bookcrossing as a photo. In conjunction with the mobile AR app this would allow the users to easily determine if the book is available by retrieving the spine image from the

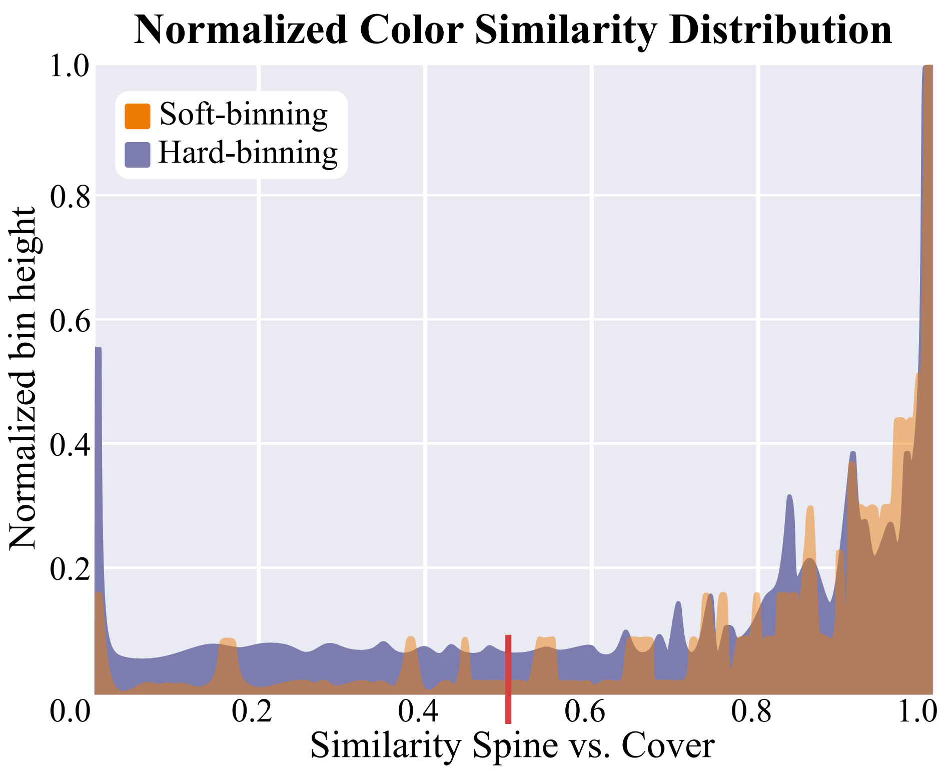

Figure 8.9: Color similarity distribution across the entire dataset of all book spines compared to the corresponding covers: A similarity of 1 means a perfect correlation between the extracted feature vectors. The red line represents the initial threshold for subsequent text recognition. In total, 21.1 % of all target books would not be considered for subsequent text recognition with non-interpolated colors (blue). By applying bilinear interpolation of the color bins 14.5 % are being rejected (orange).

web page and using it with the presented algorithm. When searching for a specific released book, there can usually only covers be retrieved from Bookcrossing using Web crawlers. Therefore we also collected all covers from our entire dataset and compared the resulting color feature vectors with the extracted features from the spines. In Figure 8.9, the red line represents the initial threshold for subsequent text recognition, with 21.1 % of target covers being incorrectly rejected due to color variations with the corresponding book spine. When bilinear interpolation [Dal05] is applied between the color bins, the false rejection rate decreases to 14.5 %.

To avoid rejecting target book spines, the algorithm must be adapted when using book covers. A color preselection cannot be carried out, since it cannot be assumed that the spine colors match those of the cover. Instead all extracted book spines are sorted in descending order according to their color similarity. Then on each spine the text recognition is performed. The weights of Equation 8.2 are adjusted without using color features in Equation 8.3. They were estimated using the same methodology as for Equation 8.2. Again, title and author are given the highest weighting from the random forest. In order to reduce computational effort, the process can optionally be aborted if a particularly high degree of similarity close to $S = 1$ is detected. Otherwise, the most similar spine can optionally be highlighted with a hint that the book also might be interesting to support serendipitous discovery.

$$S = w^T s = \begin{pmatrix} 0.7181 \\ 0.2249 \\ 0.0363 \\ 0.0207 \end{pmatrix}^T \begin{pmatrix} s_{\text{title}} \\ s_{\text{author}} \\ s_{\text{publisher}} \\ s_{\text{subtitle}} \end{pmatrix} \tag{8.3}$$

Finally, it remains to be clarified whether an accuracy of 80 % is sufficient for real world applications. It should be mentioned that users of public bookshelves do not expect perfect success in finding a book they are looking for in a particular shared library, which was also noted in the interviews. As online comments show, users of Bookcrossing are well aware that with public sharing, books may not be available anymore when they inspect the shelves in search for a specific book. In order to prevent frustration from failing to find a specific book, the design of the mobile visual search app should emphasize the possibility of serendipitous discovery of books of interest. The app should act as a recommender system and highlight relevant books.

8.9 Conclusion and Future Work

We presented a robust algorithm for detecting books in public bookcases. The prototype of the mobile AR application supports visual search in a public bookcase. Based on text or voice input, a matching book is highlighted in the viewfinder. The algorithm solves the core problem of reliably detecting books in public bookcases.

The study results show that a robust recognition of book spines is possible in real world situations. In a lab environment mobile visual search could be carried out significantly faster than unaided visual search for unsorted book shelves. The achieved accuracy of just 80 % under real-world conditions seems low, yet recognizing books under widely varying lighting conditions and in all kinds of physical states and spatial arrangements is a difficult problem.

The mobile application has to be extended to provide a wider range of functionality, according to the needs of users of public bookcases. This includes the recommendation of potentially interesting books based on a personal profile, the highlighting of books that have been added since the last visit, or even the integration with an online platform. Such an integration could be used to automatically create an online inventory of a public bookshelf, without additional administrative effort. An online inventory would help to understand in more detail what is shared and what the "shelf time" of individual books is. However, relying on an online platform may incur costs and require commitment for its long-term operation.

By also using the mobile app on one's personal bookshelf at home, recommendations could be given on which books might be interesting for others. A larger, freely available dataset could be created collaboratively, including an explicit registration of spines via the app. The app could suggest on which shelves to place specific books in order to support children, as they often cannot reach the upper shelves. An age restriction could be realized by having the app suggest placing certain books on the upper shelves.

An interesting question for future work is whether an online inventory of public bookcases destroys some of the appeal of accidentally discovering interesting books "in the wild." It is also of interest to clarify the cultural relevance of the shared books, respectively, whether the place and stock of the bookcase reflects the group of local residents and their demographics. Furthermore, it could be determined to what extent individual bookshelves are frequented, which can be deduced from the fluctuation of the stock. For example, it might be possible to find more suitable locations for new bookshelves. The analysis could be carried out automatically by taking into account location data and search queries.

The Android application and the dataset created in this work are freely available as open source software.[1]

1 https://github.com/M-Schrapel/Public-AR-Booksearch

CHAPTER 9

Interactive Acoustic Augmentation of Surfaces in the Wild

At the beginning of this thesis we motivated that floor materials in shopping malls can influence the walking speed of pedestrians [Pre73]. The perception of the floor materials is mainly influenced by the three senses vision, touch and hearing. In everyday life, there are many situations in which an adjusted walking pace could support a pedestrian. For example, an adjusted walking pace could might be used to reach a train or bus on time. In this chapter, we investigate the vision of augmented surfaces in urban areas. For this purpose, we use audio augmented realities to influence the perception of walked surfaces. Nowadays, headphones with active noise cancellation can block out external noises. This opens up the opportunity to augment footstep sounds anywhere and might change the perception of urban environments. To enable off-the-shelf augmentation of footstep sounds, this chapter presents a mobile app that recognizes footsteps based on smartphone accelerometer data. We investigate by interviews of an in-situ study how changed footstep sounds influence human perception of the environment and how augmented footstep sounds can be used as an interaction technique in the wild. Accordingly, this chapter addresses the following research questions:

> How do acoustically augmented surfaces contribute to the perception of audio augmented realities in the wild?
>
> Do augmented footstep sounds convey navigation cues?

This chapter is based on the master thesis by Janko Happe [Hap21] and a journal paper published at i-com.

- Schrapel, Maximilian and Happe, Janko and Rohs, Michael: 'EnvironZen: Immersive Soundscapes via Augmented Footstep Sounds in Urban Areas'. *i-com* (2022), vol. 21(2): pp. 219–237

9.1 Introduction

Urban soundscapes are often far from what we consider relaxing. Car engines, horns, construction sites, laughing, or clanking dishes in restaurants and other sound sources can raise environmental noise to an annoying level. Research found a relation between environmental noise and various diseases such as tinnitus, cognitive impairment, sleep disturbance, and cardiovascular effects [Isi03; Mün18; Org11]. Animals also suffer from man-made noise [Kig11], like birds singing at higher amplitude in urban areas [Bru04]. The world health organization (WHO) therefore provides guidelines prescribing a maximum noise level [Ber99; Org18]. Human ears adapt to modern environments and focus on specific sound sources from any direction while attenuating other noises around.

Nevertheless, our natural ability to filter out distracting noises is limited [Shi08]. Headphones with active noise cancellation (ANC) can effectively reduce distracting noises [Mol13] and positively influence subjective well-being while focusing on specific tasks [Mül20]. The listener can enjoy music or audio books at a lower volume [But19], but may also miss important signals like announcements on trains [Haa18]. Noise cancellation can also be beneficial for digitally supported meditation practice [Sre17]. Walking meditation is a form of focusing on one's gait by paying attention to each step [Han85]. Research has found positive effects on the treatment of depression via walking meditation [Pra14]. Typically, these exercises are performed in natural and relatively quiet environments to avoid distractions and to allow practitioners to better focus on the moment [Kab17]. For novices, it is challenging to transform a walk on a crowded and noisy sidewalk into a relaxed meditation.

Playing a soundscape through ANC headphones can present a different world acoustically, but so far there has been a lack of research on how augmented footstep sounds affect listeners in such acoustically modified worlds in urban areas. We therefore developed *EnvironZen*, a mobile app that augments the footstep sounds to immerse the user into a different world, while visual stimuli of reality are preserved. In order to investigate the influence on walking, we first examined different soundscapes and footstep sounds in an online survey and then tested the most preferred soundscapes in an in-situ field study with 16 participants. Furthermore, we investigated whether augmented footstep sounds are suitable for navigation and observed the influence of sound delays and speedups and the playback of different warning sounds, e.g. when bicycles are approaching. Repeated small interviews were conducted to capture each individual experience and perception of the urban environment and soundscapes.

9.2 Related Work

Michael Southworth showed that visually impaired people prefer other urban areas than people without visual impairment, as different characteristics gain importance [Sou69]. From this, Southworth derived the concept of the *sonic city*. R. Murray

Schafer established the term of *soundscapes* [Sch93], which is composed of three features: keynote sounds (natural noises like wind or water), signals (acoustic warnings like bells), and soundmarks (individual sounds that identify a specific location). In contrast to an acoustic environment, a soundscape is a deliberate composition of the aforementioned three features by a designer's vision [Kan16]. With the increasing awareness of acoustic environments [Kan16], cities start to actively monitor noise pollution [Bel19], as noise is related to health problems [Isi03; Org11], whereas natural sounds have been found to provide health benefits [Alv10; Ben14; Bux21].

An audio augmented reality (AAR) is a superimposition of real world soundscapes with virtual sound sources [Coh93]. It embeds computer-generated sounds into the user's acoustic environment [Gam14; Mar13]. Sonification integrates data into the sounds to convey information [Kra99]. Those acoustically embedded data can, for example, present social media messages [Din10; Her12], location-based information [Bor14], emotions [Sei17], notify specific users [But05], help to maintain a steady walking pace [Kom15; Oli06], or modulate EEG data on soundscapes to assist group walking meditation [Coc21; Coc20]. One of the first works that investigated auditory displays was Nomadic Radio [Saw00]. Notifications were embedded in the user's environment via wearable loudspeakers and voice recognition was used for input. In a mobile context, AAR is used to promote exploratory and immersive experiences as well as serendipitous discovery of public places [Bet13; Eck01; McG11; Vaz12]. For instance a leisure walk in an AAR park [Law20; Vaz12] can promote health [Org10; Pra14] and support rehabilitation from stress-related mental disorders [Cer16]. For the blind, audio output can describe the natural environment to promote nature connectedness [Ban20a; Ban20b]. Playing bird sounds through speakers in noisy urban parks can positively influence an ambience [Van20] and provide interactivity, e.g. via smart clothing [Kob08]. Following the concept of a *Sound Garden*, anyone can place an auditory virtual park at a specific location to offer listeners a new experience [She06; She07] or guide tourists between landmarks [Bed95; Bol18; McG09]. This concept can also be adapted to recreate the soundscape of historical places [Sik18] or to represent historical events [Rei05], e.g. to enhance touristic experiences [Cho21] and museum guides [Zim08] or stories [Ros16]. Thereby, spatial audio can support the design process of a soundscape [Hon17] and 3D sound can contribute to achieving immersion in audio-augmented worlds, e.g. for sound-based games in urban areas [Cha16]. Furthermore, spatial audio can be used for guidance. For example, the direction from which music is played [Alb16; Hel20], or individual instruments [Hel18] can be used for guidance. To determine the orientation of the user in relation to the played spatial sounds, sensors in the headphones or the smartphone can be used as a virtual directional microphone [Hel15; Hel14]. Cocooning refers to both active and passive acoustic isolation [Bij10]. It involves suppressing one's own sounds, such as breathing and footsteps, as well as external sound sources. Nowadays, Hearables

[Cru19; Pla18], Earables [Kaw18] and headphones with active noise cancellation allow for acoustic transparency to let their listeners stay aware of the environment. McGill et al. envisioned that in the future users will immerse themselves into an acoustically interactive space with personal and private auditory displays [McG20]. They investigated the impact of acoustic transparency with a focus on external sound sources and found that acoustically transparent and isolated headsets have an effect on the realness and presence of the content.

In virtual realities, footstep sounds enhance the perceived level of presence [Ker20] and also affect walking behavior [Hop19; Ser11; Wil20]. In the context of AAR, sonified footsteps offer a certain degree of realism and interactivity with virtual acoustic worlds. Concepts of mounted step sensing and footstep playback on urban running routes aim to provide a more natural running experience [Ren21]. Vibration feedback from shoes was used to change the perceived physical structure of a surface [Son18; Tak10]. Different natural surface simulations by footstep sounds were found to influence the biomechanics of walking [Tur15], perceived body weight [Taj15], emotions [Ley19], step length [Cor20], and walking pace [Men10; Tur13b]. Those effects were also used to manage gait disturbances in Parkinson's disease [Rod14], gait rehabilitation in chronic stroke patients by pitched natural footstep sounds [Gom20], and foot pressure balance for runners [Ohn19]. Music that was modified by footsteps showed no perceived effects on user's gait [Haj16]. Using non-naturalistic footstep sounds showed no significant effects on walking patterns [Bre10]. The preferred volume of footstep playback was found not to significantly differ from the volume of the soundscape [Tur13a]. Most of the mentioned studies that analyzed augmented footsteps were carried out under controlled lab settings and did not involve users walking in real application scenarios. Hence, the results are often based on small distances. Furthermore, effects due to the measuring equipment, e.g. shoe prototypes, could influence the results.

The existing body of related works shows that soundscapes and footsteps contribute to the immersion in virtual worlds [Ker20], in which footstep sounds also influence walking behavior [Hop19; Ley19; Men10; Ser11; Taj15; Tur15; Tur13b; Wil20]. There is a positive influence of natural sounds [Alv10; Ben14; Bux21; Cer16] on health, and augmented footsteps sounds may serve as an unobtrusive feedback channel [Gom20; Ohn19; Rod14]. In addition, soundscapes themselves can provide special cues about the environment [Bor14; But05; Din10; Her12; Sei17]. However, so far the main focus of deploying interactive soundscapes has been on city parks [Ban20b; Law20; She06; Van20; Vaz12], guidance with soundscapes [Ban20b; McG09; Rei05; Sik18; Zim08] and virtual reality with augmented footsteps [Hop19; Ker20; Ser11; Wil20]. Our vision is motivated by noisy and stressful city center environments [Bar17] that can now be masked out by ANC headphones [But19; Mol13; Mül20]. This might enable the transformation of any acoustic environment into a natural soundscape and provide interactivity in AAR via augmented footstep sounds. Relaxing and exploratory experiences to promote health by walking [Org10; Pra14; Vaz12] thus can be implemented at any location and extend the capabilities of sound gardens [She06]. The created natural soundscapes can be enhanced by warning sounds that fit into the virtual acoustic environment in order to notify the listeners [But05] of hazardous situations, without completely taking the user out of the immersive experience. For instance, approaching vehicles that share their location in real time can communicate with pedestrians in virtual acoustic worlds [Cha17a]. However, in previous works, external sounds from the virtual environment were used as cues, while virtual sounds produced by the listeners themselves, e.g., footsteps, were not considered. For example, when accidentally leaving a guided audio path [McG09], special footstep sounds can notify users that they walk the wrong way, as they would when walking beside a gravel path in the forest. Accelerated or delayed footstep sounds could unobtrusively hint users to hurry up, e.g. to catch a bus, or calm down to be more relaxed [Men10]. By switching between different soundscapes, users can experience a multifaceted path with a rich variety of impressions [Tur13b]. In this chapter, we explore the vision of augmented interactive soundscapes in city centers with augmented footstep sounds and its influence on listeners. In contrast to [McG20], we focus on isolated auditory displays and investigate the impact of footstep sounds and their capabilities in audio augmented realities. As a research vehicle we use the *EnvironZen* application. To ensure universal and ubiquitous usage, only smartphone acceleration data are applied to establish real time footstep detection and sonification. This approach does not require instrumenting users with additional equipment that may have an impact on the natural user's gait. The main contributions of this chapter include (1) an open-source application for augmenting footstep sounds, (2) an evaluation of individual perceptions of urban stimuli in audio augmented realities, (3) the sonification of augmented footstep sounds to convey navigation cues, and (4) an evaluation of the effects of accelerated and delayed footstep sound feedback.

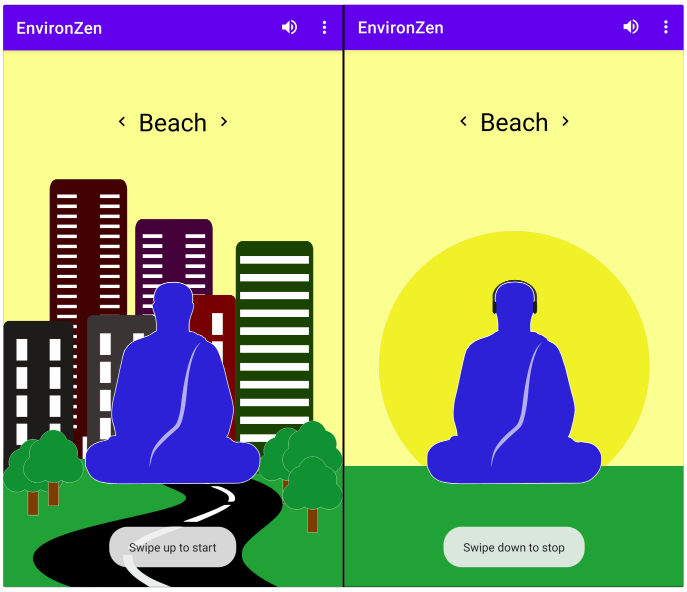

Figure 9.1: Left: Start screen of EnvironZen with inactive soundscape; the urban environment signals the unmodified urban sound environment. Right: Swiping up activates the "Beach" soundscape; swiping left or right selects other soundscapes; swiping down stops the audio augmentation.

9.3 EnvironZen Application

EnvironZen is designed to transform any acoustic environment into a relaxing soundscape. Based on motion data of the most ubiquitous device – the smartphone – [Bal06] the application aims to recognize and augment footsteps, anywhere and without any additional equipment. The name is composed of *Environment* and *Zen*. Although we do not fully relate our app to the principles of *Zen Buddhism* and walking meditation [Han85; Kab82; Kab13], we aim to encourage users to have a relaxing, stress-relieving walking experience. The idea behind EnvironZen is to motivate a walk, e.g. when going to work or during a lunch break in city centers, where due to the short time of the break no parks are within reach. Since visual stimuli are maintained, the approach is compatible with busy environments, like an urban sidewalk. The auditory footstep augmentation was developed to ideally immerse the user more fully in the auditory world. The state of the Buddha figure in the startup screen of EnvironZen (Figure 9.1, left) displays the app status. If the Buddha figure is shown with an urban environment in the background, the sound augmentation is off. Swiping up changes the figure: the Buddha now wears headphones and the urban background fades away, being replaced by a circular "halo". The selected soundscape starts playing ("Beach" in Figure 9.1). Swiping left and right changes the soundscape and footstep sounds. By the background color and the name of the soundscape, users can identify their selection. Footstep sounds are played only when the smartphone is put in the trousers pocket, which is recognized by the proximity sensor of the display. In the menu settings users can also disable the pocket mode. However, EnvironZen's algorithm was optimized for carrying the smartphone in the pocket since the visual perception of the environment is to be preserved. The user can adjust the volume of the footstep sounds and of the soundscape via the speaker button on the top of the screen. Additionally, in the menu users can choose a navigation mode to be guided to a selected location. In this mode, the footstep sound is changed to a wooden step sound when users are walking in the wrong direction. Further, a story mode was implemented that switches randomly between the soundscapes every three minutes, in order to tell the listener a story without any spoken words. The term *story mode* is intended to encourage the user to engage with the virtual acoustic environment. While switching, a ten-second crossfade between the soundscapes and the footstep sounds ensures a pleasant adaptation for the listener. For study purposes, the app can be connected to another smartphone via Bluetooth to start recordings and to activate various study parts. The application was created for Android 6 or higher and is freely available at GitHub [1] along with the audio clips.

9.3.1 Step Detection

Step detection algorithms can analyze motion measurements in either the frequency domain [Kan18] or in the time domain. Time domain approaches include, but are not limited to, thresholding [Alz12], zero-crossing counts [Jay13], and peak detection [Yin07]. Our step detection algorithm is based on the recommendations by Brajdic and Harle [Bra13]. However, we had to extend and reconfigure the proposed windowed peak detection algorithm [Bra13] in order to support real-time step counting. In contrast to Brajdic and Harle, who used a fixed window size of 590 ms for peak detection, we use a window size of 500 ms and change it to 330 ms when the acceleration intensity is above a threshold. We included this adjustment to support activities like short runs, e.g. to catch a bus, for which a fixed window size showed highly inaccurate step counts. Thus, the algorithm covers two states: walking and running.

In Figure 9.2, *MovingAverage(duration)* is an object that computes the moving average over the last *duration* seconds (window size). The *add* method adds another acceleration magnitude value. The *avg* method returns the current moving average. The magnitude is calculated as $accMagn = \sqrt{a_x^2 + a_y^2 + a_z^2} - g$. The earth's gravity ($g = 9.81\ m/s^2$) is subtracted in order to zero the result when no activity is performed. The magnitude is then filtered with a moving average over 310 ms, as proposed by [Bra13]. The parameters p_1 and p_2 indicate window sizes (in seconds), while p_3 to

1 https://github.com/M-Schrapel/EnvironZen

p_5 serve as thresholds to detect signal peaks. The parameters have been optimized by an evolutionary algorithm using a Gaussian mutation operator. The window size *minWindow* is initialized with 500 ms and is updated for each measurement depending on the intensity filter. In case the intensity filter exceeds the threshold p_3, *minWindow* is set to 330 ms to allow for accurate step detection while running. The assumption here is that the acceleration is higher for running than for walking. If the time since the last recognized step is less than *minWindow*, then no step is reported. A step is recognized when the filtered magnitude *accMagFiltered* is above the threshold *thresholdFilter.avg* multiplied by the parameter p_4. Further, the filtered magnitude *accMagFiltered* must be above the absolute threshold p_5 to detect peaks in the signal.

```
minWindow = 0.500 sec
magFilter = new MovingAverage(0.310 sec)
thresholdFilter = new MovingAverage(p1 sec)
intensityFilter = new MovingAverage(p2 sec)
procedure DetectStep(accMag)
    magFilter.add(accMag)
    accMagFiltered = magFilter.avg
    thresholdFilter.add(accMagFiltered)
    intensityFilter.add(accMagFiltered)
    if intensityFilter.avg > p3 then
        minWindow = 0.330 sec
    else
        minWindow = 0.500 sec
    end if
    if timeSinceLastStep > minWindow then
        if accMagFiltered > thresholdFilter.avg * p4 and accMagFiltered > p5 then
            return true
        end if
    end if
    return false
end procedure
```

Figure 9.2: Step Detection Algorithm.

To optimize this algorithm, two experimenters (25 years, 75 kg, 1.80 m and 32 years, 97 kg, 1.89 m) recorded accelerometer measurements with a mobile phone (a *Xiaomi 9T* and a *Huawei Mate 20 Pro*, respectively) placed in the front left trousers pocket. They walked for five minutes on sidewalks to record accelerometer data. One developer (32 years) is an experienced runner, recording additional data at an almost constant pace of 10 km/h while running in bear-foot shoes. A total of 230 steps were extracted for the optimization of our algorithm. Samples were discarded for which the magnitude's signal peak did not exactly indicate a heel strike.

The dataset of 230 steps was used to optimize the parameters p_1 to p_5 with a fitness function that runs the step detection algorithm over the entire dataset until all steps are recognized. The algorithm is provided with the actual number of steps and has to locate them in the dataset. In each iteration, all parameters are mutated with a Gaussian function. The steps are then computed and the new result is compared to the previous result. For this purpose, the *(1+1)* evolutionary algorithm [Dro02] selects the previous (parent) or new parameterization (child) with number of recognized steps closest to the real number of steps, i.e., the lowest fitness score, and continues the mutation and calculation until no steps are unrecognized. The resulting parameters are shown in Table 9.1. The algorithm was subsequently tested by one female and four male volunteers (age: M=31.6, SD=13.2 years; height: M=1.77, SD=0.1 m; weight: M=77.6, SD=8.45 kg). They installed a test application on their own smartphone (two *Huawei One Plus 6*, *Samsung Galaxy S5* and *S10e*, *Google Pixel 4a*) and were introduced to walk for five minutes on a pedestrian walkway with the phone in their left trousers pocket and headphones on. During the test they heard a beep sound every time a step was detected. Although two participants mistakenly put the phone in their right trousers pocket, all testers reported an accurate and instant step detection when their feed touched the ground. To evaluate the performance of the step sensing algorithm, we have built a pair of sensing sandals with attached force sensing resistors on the sole. An Arduino measured both the pressure on the sole and the audio signal amplifier output from a *Zealot H6* Bluetooth headset. The headset is connected to a *Xiaomi 9T* smartphone running our application. We recorded timestamps, when the sole pressure changed and a beep sound is played. On the phone we record the time from a triggered accelerometer sensor event to immediately before a footstep sound is played. One author recorded time measurements from

Table 9.1: Algorithm parameterization based on 230 recorded steps from two people during running and walking. The parameters were optimized with an evolutionary algorithm using a Gaussian mutation operator.

	Values	
Parameter	*Initial*	*Optimized*
p_1	0.2	0.629
p_2	1.5	1.365
p_3	4	3.893
p_4	1	1.017
p_5	0.7	0.715

30 steps walking along a corridor. We measured that the mean time from the foot touching the ground to sound being emitted is 351.9 ms (SD = 104.8 ms), of which on average 300 ms are due to the Bluetooth delay [Tri21]. The average step sound delay of 51.9 ms includes an average of 3.9 ms (SD=1.4 ms) from the moment the sensor routine detects a step to triggering a sound event on the smartphone.

9.3.2 Soundscapes & Footstep Sounds

We use different royalty-free soundscapes and footstep sounds from the web for our application. Ten different, randomly selected footstep sound variations for each surface ensure that listeners get the feeling of uniqueness for every step they take. Furthermore, the measured duration between two detected heel strikes is used to model the volume of the footstep sound in order to reinforce this impression. The spectrograms shown in Figure 9.3 visualize some of the implemented soundscapes and their frequency distributions. *Beach* is composed of an ocean wave keynote sound, which is accompanied by seagull cries as soundmarks. In this soundscape, users walk on sand to create the impression of a walk on the beach. The spectrogram shows a relatively uniform frequency distribution, which is primarily defined by the ocean wave peaks. *Creek* is defined only by the keynote of flowing water. Although the spectrum is similar to *Beach*, there are no clear soundmarks here and the listener walks on deep puddles. *Forest* is a rich soundscape with singing birds as soundmarks and a small creek as a keynote sound. The listener walks on dry leaves through this soundscape. The spectrum here is characterized by high-frequency birdsong. Similarly, *Jungle* takes the listener to the soundscape of a rainforest. Besides singing birds, chirping crickets and frogs can be heard. The reverb effect creates a wide and extensive atmosphere, in which the listener walks on the same dry-leaves-sound as before in the forest. The abundance of the environment can also be seen in the high intensities in the lower frequency range. The *Mountain* soundscape represents an empty place, only defined by blowing wind and a screaming eagle, which is clearly identifiable in the spectrum by the two peaks at 5 and 30 seconds. Here, the listener walks on little stones or gravel whereby the strong wind hints at the location. Likewise, the *Snow* soundscape is defined by cold, blowing wind, which presents the soundscape much more dominantly than the wind in the mountain soundscape. Comparing the spectra of *Snow* and *Mountain* reveals this difference. Here, the listener walks on deep snow through this rough and cold environment. Further, we implemented a *Rain* soundscape, in which continuous rain and a footstep sound of walking in a puddle creates the acoustic environment. The *Musical* soundscape was added as a playful, interactive music environment, in which the act of walking generates music similar to a floor piano. A continuous drum beat is played and to every footstep, a random note of a kalimba is played. In this soundscape, the effect of the footstep volume adjustment based on the duration between two steps becomes particularly useful for creating unique sounds with every step.

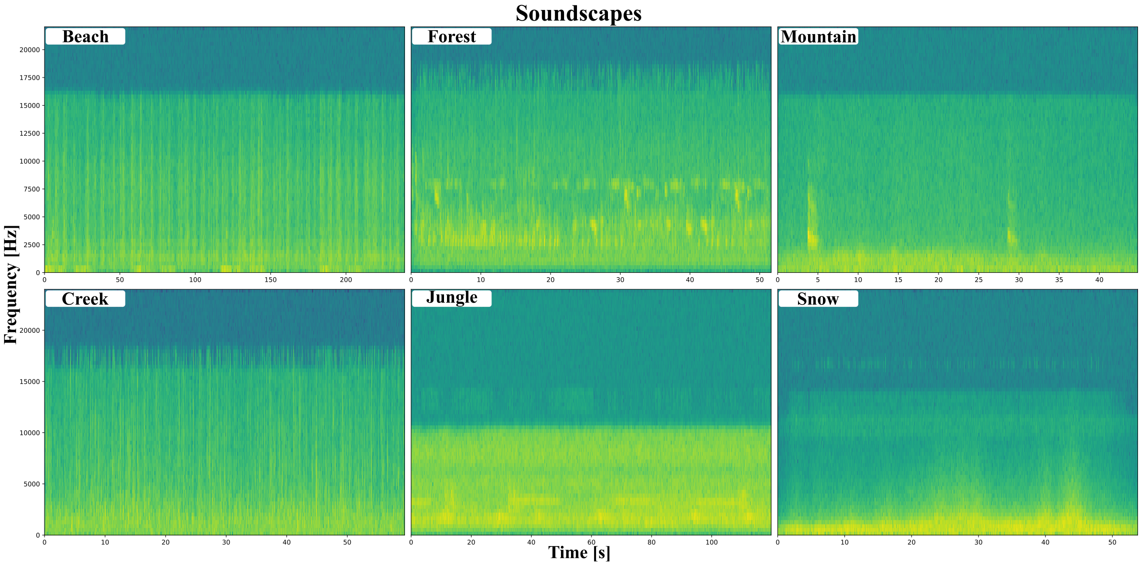

Figure 9.3: Spectrograms of the various soundscapes: During a walk the background sounds are played continuously.

9.4 Online Survey

We aimed to evaluate our soundscapes and the footstep sounds in order to find the most preferred and pleasant sounds for a subsequent in-situ field study. In addition, we wanted to investigate which warning signals that match the soundscape can also be used to interpret approaching vehicles, such as scooters or bicycles. Sixteen individuals (four female and twelve male) aged between 22 and 32 years (M=25.2, SD=2.3 years) were surveyed. They were asked about their headphone usage and whether they use noise canceling headphones. Further, we added questions regarding exercises for relaxation purposes with and without digital support through wearables or apps.

After the first set of demographic questions and headphone usage, respondents were introduced to put on their headphones in a quiet environment and to listen carefully to the subsequent audio samples. In the first part, they had to listen and rate the footstep sounds from Section 9.3.2 on a 5-point scale from very unpleasant to very pleasant. The survey continued with a vision video of our application, in which we introduced the idea of augmented footstep sounds in urban areas. As before, the respondents listened to and rated each soundscape including the footstep sounds. The last part showed a 10-second video of a walking person and an approaching bicycle to demonstrate the idea of soundscape-related warning sounds. The audio track consisted of a soundscape and footstep sounds from Section 9.3.2. In this video, the approaching cyclist is shown ringing the bell, but the bell sound is replaced by different warning sounds, depending on the soundscape. The warning sounds were a galloping horse for *Beach*, a woodpecker for *Forest*, sleigh bells for *Snow*, the roar sound of a tiger for *Jungle*, and an electric guitar for *Musical*. After each warning sound, we asked the participants to judge how well the sound fits into the scenery and whether the sound is suitable to warn of approaching vehicles. The soundscape *Mountain* was not surveyed on warning sounds, as we here use the bicycle bell as the warning sound in the in-situ study.

9.4.1 Results

Eight participants stated to regularly use headphones while walking and doing outdoor sporting activities. Seven of them use ANC headphones for listening to music in public places, during transportation and at work. Six participants reported to perform meditation exercises like yoga or breathing exercises for relaxation purposes. Four of them support their exercises with mobile applications. Eight stated that they have never used meditation applications for recreational activities. Six of them declared that they had never done meditation exercises before.

Figure 9.4 depicts the ratings of the sounds presented in the survey. *Forest* was found to be the most relaxing presented soundscape, followed by *Beach*, and *Mountain*. A Friedman test indicated that our soundscapes were rated differently $\mathcal{X}^2(6) = 24.234, p < 0.001$. A subsequent post-hoc Tukey test showed that *Forest* was rated significantly higher, while *Musical* was rated lowest among the others. The respondents could not imagine using *Musical* for recreational purposes, but found the *Guitar* to be the most appropriate warning sound, followed by the *Sleigh Bell* and the *Tiger* sound. In the *Forest* soundscape, the *Woodpecker* was appropriate but less dominant as warning sound. The *Horse* at the *Beach* was mentioned to grab attention, but our respondents found the sound inappropriate on the beach. Thus the sound was changed for the in-situ study to a *Ship Horn*, introduced and

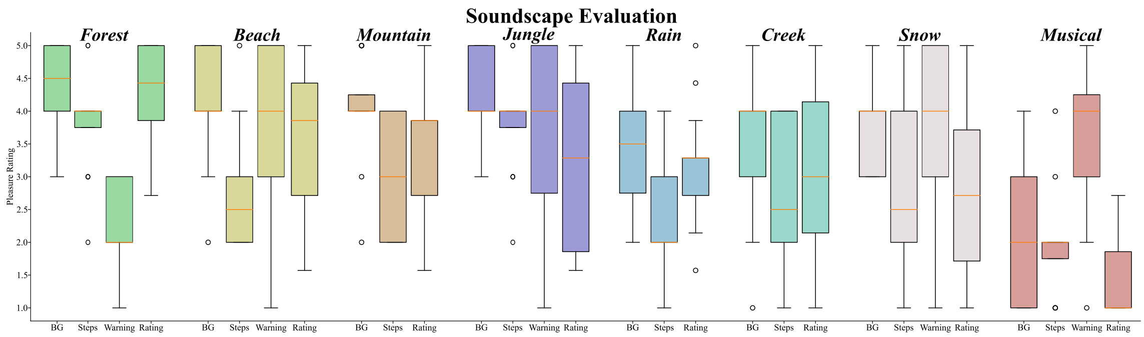

Figure 9.4: Ratings of the implemented soundscapes, sorted from highest to lowest score. A score of 5 presents the highest rating. The pictograms in the bars indicate the footstep sounds. BG is the background environment without any augmentation (just ANC). The first three soundscapes were also evaluated in the in-situ study. The colors of the bars separate the soundscapes.

rated in the opening questionnaire of the in-study. *Leaves* was found to be the most suitable footstep sound for *Forest* and *Jungle*, followed by *Gravel* on the *Mountain* and *Sand* on the *Beach*. *Leaves* was remarked as similar sound to a plastic foil by two respondents. In the *Snow* soundscape participants perceived the footstep sounds on *snow* similar to a wooden floor. From the findings we selected *Forest*, *Beach*, and *Mountain* for the in-situ study. Since the galloping *Horse* in the *Beach* soundscape was rated low, we decided to change that sound to a *Ship Horn* for the main study. Further we asked which other soundscapes they would like to immerse in. Participants mentioned a drip stone cave, the soundscape of the moon, a medieval township, a sunny meadow, New York City at night, and one wanted to have a trampoline sound for the footsteps in a fun park.

9.5 In-Situ Field Study

We conducted a field study to find out how augmented footstep sounds can be used in urban environments. We recruited sixteen participants (three women and 13 men) aged 21 to 58 years (M=28.9, SD=9.7 years) for the in-situ study. Seven of these participants (M=26.9, SD=2.4 years) also participated in the previous online survey. One day before each volunteer participated in the study they were kindly instructed to fill out an initial questionnaire. We asked the same demographic questions as before about headphone usage and introduced the topic of the survey. In addition, we repeated the warning sound test for two selected soundscapes *Forest* and *Beach*. The intention was to increase awareness of the warning sounds for the following study so that the participants are able to interpret these signals in the soundscapes. For the *Beach* soundscape, two sounds were presented. At first the galloping horse sound from the online survey and second a ship horn sound. The latter was chosen for the study as this dominant sound can be more easily associated with a marine environment.

The study took place on weekdays between noon and 5 PM in the city center of Hannover, Germany on a crowed pedestrian sidewalk next to a city street. Two construction sites, cars, restaurants, and stores shaped the lively acoustic environment and thus gave us the opportunity to test our application in the proposed scenario. Figure 9.5 depicts the study route taken in summer season of 2021 mainly during sunny weather with some wind blowing on several days. At the starting point, the experimenter handed out the *EnvironZen* App installed on a *Huawei Mate 20 Pro* to the participant. During the study the experimenter remotely controlled the app via Bluetooth with a *Huawei OnePlus 3* smartphone. The participant was instructed to comply with the traffic rules and asked put on the *Sony WH-3000-XM3* ANC headphones.

After calibrating the ANC algorithm to the participant's ears, the participant adjusted the volume of the soundscape and the step sounds to their preferred level

Figure 9.5: Image at starting point (right) and study route map (left): at route segment A participants walked with headphones in ANC mode without any sounds. From B to D the soundscapes *Beach*, *Forest*, and *Mountain*. Augmented footstep sounds *Sand*, *Leaves*, and *Gravel* were played in the sections marked in green. Segment E was used for storytelling with footstep sounds and for playback of warning signals when vehicles are approaching. The pictograms indicate the soundscapes and the yellow marks show construction sites. Segment F was used for navigation. G and H were used for delayed or accelerated footstep sounds using the *Forest* sounds.

and put the phone in their left trouser pocket. To make them familiar with ANC, they walked route segment A with ANC on, but without any sound augmentation to the end of the first sidewalk. The experimenter followed the participant all the time at a short distance. At the end of the route segment the participants took off the headphones and the experimenter interviewed them about the impression of the noise cancellation. All interviews were audio recorded for later analysis. Afterwards on the route segments B, C, and D the three soundscapes *Beach*, *Forest*, and *Mountain* were played in counterbalanced order. Until the middle of each route (marked in blue in Figure 9.5) the participants listened to the pure soundscape without footstep augmentation. Then (marked in green in Figure 9.5), the augmented footstep sounds *Sand*, *Leaves*, or *Gravel* were played depending on the soundscape. At the end of each route segment participants were interviewed about their impression on the temporal accuracy of augmented footstep sounds, whether they experienced immersion into the soundscapes, what they remembered to having heard, and whether and in what way the footstep sounds contributed to the scenery. When participants remarked that the volume of the soundscapes or of the footstep sounds was not appropriate, they had the opportunity to readjust the volume. Furthermore, we recorded GPS and accelerometer data, as well as timestamps, volume, and recognized steps.

Route segment E is divided into the three soundscapes *Beach*, *Forest* and *Mountain* with augmented footsteps as shown in Figure 9.5. This segment was intended to evaluate whether footstep sounds can contribute to more immersive storytelling and how warning sounds are perceived that aim to direct the user's attention to approaching vehicles. The warning sounds are realized with spatial audio in eight compass directions. The intended story was simple: A walk begins on a beach, then leads though a forest, and finally ascends to the top of a mountain. When vehicles like bicycles or e-scooters were approaching, the experimenter sent a command to the participant's smartphone that specified the direction of the vehicle and the soundscape-related warning sound. The warning sounds *Ship horn*, *Woodpecker*, or *Bicycle Bell* were played from the vehicle's direction in relation to the user. This does not provide a controlled study setting, but it does make the participant more aware of approaching vehicles while using ANC headphones for their safety. The sound was linearly cross-faded between two adjacent soundscapes to avoid hard transitions. Before this part began, participants were only told that they were about to hear a story and warning sounds when vehicles are approaching to which they should pay attention to. Afterwards, participants were interviewed about the story and warning sounds they heard.

Route F was used to evaluate whether footstep sounds can be employed for navigation purposes. In the *Forest* soundscape participants walked on *Leaves* as long as they walked in the right direction. When they walked on the wrong path, the experimenter remotely changed the footstep sound to *Wooden Planks* with his phone. Before this part began, participants were informed that the system would navigate them a destination unknown to them, which they can identify by a *Gong* sound. Further, they were informed that they would walk in the wrong direction when they hear themselves walking on a wooden surface. When reaching the location, participants were asked how they experienced the feedback.

The aim of the last part on route segments G and H was to measure the influence of asynchronous footstep sounds. Again the *Forest* soundscape was used together with the *Leaves* footstep sound. The participants were introduced to walk to the end of each route without giving them any information on the intention of this study part. By averaging the time between the first five recognized steps we measured the current walking speed. Based on this calculated average step duration a footstep rhythm continued that, depended on the study part, ether increased or decreased by five milliseconds every two seconds. After 60 seconds, or the initial averaged step duration plus or minus 150 milliseconds, the footstep playback continued with this maximum step duration offset. Each participant either began by hearing increasing or decreasing footstep sound periods on route segment G and then continued with the other method on route segment H. After each segment, participants were interviewed if they noticed anything. Afterwards, participants were handed out a final questionnaire for gathering feedback about this study and the presented technology. We asked which study part they liked most, which soundscape they preferred and why, how accurate they found the timing of the step sounds to be, whether step sounds provided a more immersive and relaxing experience, as well as in which context they would use such audio augmentations. Furthermore, the participants were given the opportunity to give us free-form feedback.

9.6 Evaluation

In the following, the results of the in-situ field study are presented.

9.6.0.1 Augmented Footsteps vs. No Augmentation

Initially, the participants walked with active noise canceling to familiarize themselves with the technology. In the subsequent interview eleven participants mentioned that it was unusual but pleasant to block the environmental noise. One noted that he never before realized how noisy the environment actually is. Two remarked that the headphones amplified their own footstep sounds while walking. In the following, results of our interviews from the counterbalanced soundscapes are presented.

In the *Beach* soundscape, participants mostly mentioned the ocean wave sound first as a dominant keynote sound. One participant thought he mistook the sound of waves for approaching cars, and another noted that he had to look around to see where the water was coming from. P2 stated *"It was confusing, it felt like I'm walking on a beach but I couldn't see it."*. Similarly P8 said *"It was surreal to walk on the lively street but hear the soundscape of a beach."*. Interestingly, some participants incorporated their visual impressions into the soundscape and felt even more immersed in a beach environment. P1 and P11 remarked that they had the impression of walking on a beach promenade, because of the restaurants nearby. P10, P12, and P14 had been on a beach vacation shortly before and felt taken back by the soundscape. With footstep augmentation, P6 noted *"With the step sound it was different, I was more focused on the acoustic environment, it was more intense."*. However, P16 perceived, in contrast to the other participants, the steps as not being in sync with his footsteps and felt like someone is walking behind, which he found confusing. P10 remarked that augmented footsteps suppressed the sound of his own footsteps, which made the soundscape more immersive. P11 mentioned that the augmented footsteps reinforced the soundscape and gave a more intense feeling, making him focus more on his own footsteps and breathing. P8, P9, P12, P14 and P16 found the footstep sounds slightly too loud, which made it more difficult for them to immerse themselves in the environment. P12 mentioned that he was able to get used to the volume after taking a few steps. The participants were therefore given the opportunity to readjust the footstep volume for the following parts.

When the experimenter asked after the *Forest* soundscape what kind of soundscape they heard, seven participants mentioned the *Leaves* step sound as a remarkable environmental feature. P15 rather mentioned dry grass and perceived the environment, like P10 and P16, in the very deep forest. P1 found the environment with step sounds more immersive but remarked *"The footstep sounds came a bit too early. But when I began to imagine dry leaves on the ground making sounds before the feet touches the ground, I felt even more immersed in the forest."*. P9 mentioned that the footstep sound matched his gait making the soundscape more immersive. For P5 it was challenging to concentrate on the soundscape due to the wind during the study. A misaligned footstep sound playback was perceived as disturbing. P6 repeated the perceived effect of focusing as in the *Beach* soundscape. For P4, the footstep sound increased the presence and the previously decreased footstep volume made it more realistic. P13 remarked that *"the step sound was too beautiful to be realistic."*. For P10, P14 and P15 the step volume was perceived as too loud and had to be adjusted. P12 stated that he immersed more in the soundscape by looking at the trees on the side of the road. Similarly, P3 reported that he focused more on his visual perception.

The soundscape *Mountain* was perceived very differently among the participants. P3 imagined to walk in a forest on puddles which was less pleasant to him. P7 interpreted wind sounds as ocean waves and found the footstep volume too loud, which was readjusted in consequence. P10 found himself in a primeval forest with a screaming parrot. During the interview, when mentioning the walked stony path, he noticed that the soundscape fits better to a mountain after all. P11 could not distinguish whether the wind noise was coming from cars on the road or the soundscape. P14 found the footstep sounds less synchronized with his gait. In contrast, P1 liked the quiet sounds. Although he perceived the soundscape as less immersive, he favored the footstep sound. Similarly, P15 mentioned *"I could almost feel the surface structure of the ground."*. P2 favored the empty, quiet soundscape. The footsteps on the dry path and the wind were perceived as very relaxing. Readjusting the footstep sounds was remarked as more pleasant for P13. P12 associated the footsteps with a swamp. He remarked that in his bear-foot shoes he felt on the side way with the footstep sounds like he would *"step into something."*. Thus, he decided to shortly walk on a small grass field beside the road which was more pleasant to him. P6 used the quiet soundscape to focus on his gait while walking. Most interestingly P16 smiled and said *"I felt taken back to my childhood. The soundscape matched the place where I grew up. I walked again on the stony path next to corn fields and a forest on the other side."*. Further he continued *"Without the footstep sounds I couldn't have immersed myself back."*.

9.6.0.2 Storytelling and Warning Sounds

On the story path back to the starting point warning sounds were interpreted correctly, but none of the participants really integrated the approaching vehicles in their imagined story. P1 found the composition made of *Beach, Forest, Mountain* like a short walk through Japan. Beginning from the sea he walked up to the mountains.

He remarked the *Ship Horn* as very appropriate to catch his attention. P2 interpreted the ship as a train and could not imagine where it was coming from in the *Beach* soundscape story. The windy weather on this day contributed to immersing into the entire story. Although the *Bicycle Bell* was easy to interpret in the *Mountain* soundscape, the participant was less aware of the traffic and instead felt more relaxed. P2 also stated *"I would like to have control when warning sounds are played. When a bicycle was approaching in front of me I could already see it."*. P3 has perceived the story of a plane crash on a lonely island. He remarked that the *Woodpecker* sound was less dominant and not appropriate for warnings. He watched the traffic but mentioned that he could also immerse into the story. Similarly, P4 stranded on a lonely island. P5 made a travel around Europe starting from the Northsea to the Alps. P6 wandered to a river source and completely forgot the study situation. The crossfade of the soundscapes was remarked as *"extremely convincing."*. P7 felt teleported through different worlds and almost forgot the city environment. The *Woodpecker* warning sound was remarked as *"surprisingly suitable and fitting into the soundscape."*. P8 found the *Bicycle Bell* funny and surprising. It fitted well into the overall reality. P9 made a relaxing day on the beach and went back home in the mountains. Similarly, P10 imagined a tropical island and returned to the resort. He even felt that he could concentrate better on the traffic. P11 felt remembered of a full vacation. P12 made a walk to mountains but had the feeling he could concentrate less on the traffic. P13, P14 and P16 made similar walks. P15 stranded after shipwreck on a lonely island in search for food. The warning sounds did not match his story, which made him feel taken out of his story.

9.6.0.3 Navigation

Participants were asked how synchronous they perceive the footstep sounds now after using the technology for a while. In 8 cases (P1, P2, P6, P12, P13, P14, P15 and P16) participants remarked that the step sound is now more synchronous as it was in the beginning. P5 was the only one who still perceived the footstep sounds as asynchronous with slight improvements. P3, P7 and P14 were impressed about the synchronous footstep feedback. P8 commented that the synchronicity of the footsteps depends on the sound and its dynamics. P16 remarked that he had to get used to the footstep sounds in the first five minutes of the study. Afterwards, he perceived the sounds and his footsteps as part of the soundscapes. The intended navigation path was walked by two people without taking the wrong path. Eleven participants stated that they instantly understood the footstep feedback on the wrong path and changed the direction immediately. P12 commented that this kind of feedback has a *"binary fashion"* meaning that when there are more than two possible directions, he had to try the other paths. P13 and P16 also mentioned that. P11 would like to hear the feedback left or right to identify the right way. When P2 tried the directions she felt watched by strangers.

9.6.0.4 Asynchronous Footstep Playback

In 14 cases, people noticed the faster footstep playback. P5 found the step sounds to be asynchronous over the entire study and couldn't identify any changes. P8 could also not perceive any differences although he previously found the step sound to be mostly synchronous. Five participants perceived the faster feedback as stressing, five felt pushed by the footstep sounds and four remarked the feedback as unpleasant. P14 mentioned a feeling of walking beside a person or dog. P16 noticed that he started to walk faster and then decided to slow down his pace. P15 walked his own pace but felt like walking faster.

In contrast, nine participants found no difference when the footstep sounds were played slower. Again, P5 didn't perceived any differences. P10 and P11 thought the slower pace might be an imagination. Five participants (P1, P2, P3, P4, P11) found the slower pace to be relaxing or pleasant. Four participants (P2, P9, P11, P13) reported an urge to slow their pace. P14 again reported the feeling of walking beside another person. He also perceived the walked surface as *muddy*. P2 was persuaded by the footstep sounds and stated afterwards: *"When I noticed the slower footstep rhythm I realized that I really walk too fast and should slow down myself to be more relaxed."*.

In Figure 9.6 the influence of the playback is visualized. A Mann-Kendall test was performed for each study part. We could not identify a significant trend when footsteps were played faster ($p = 0.79, z = 0.27, \tau = -0.04, s = -17.0$). When footstep sounds were played slower, a Mann-Kendall Test identified a significant increasing trend of the step durations ($p = 0.0072, z = 2.69, \tau = 0.34, s = 0.0007$). From the Figure 9.6, the trends can also be identified.

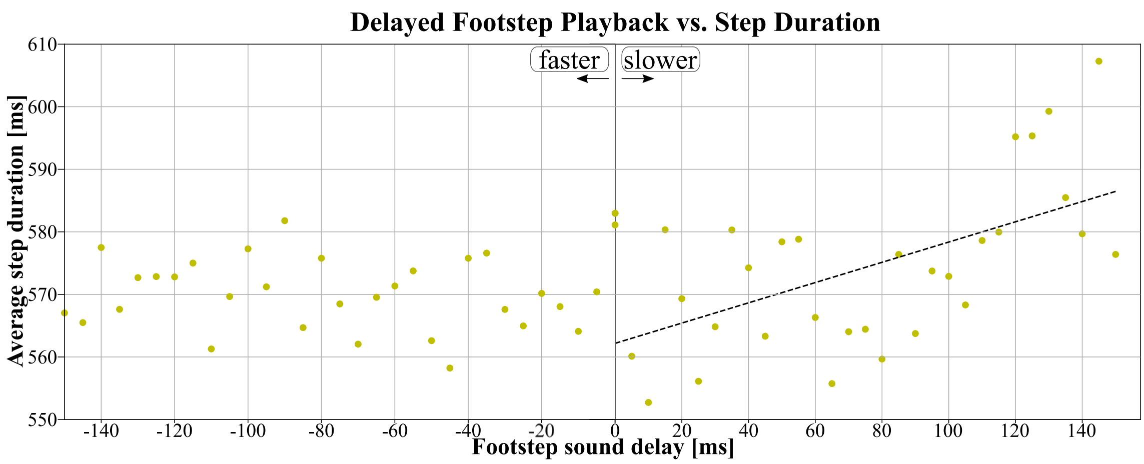

Figure 9.6: Footstep playback delay vs. step duration: A negative delay corresponds to a faster walking speed. The dashed line visualizes the significant trend found by a Mann-Kendall test. By decreasing the step duration we could not identify trends.

9.7 Discussion

Soundscapes do not necessarily have to compose a rich sound environment to be perceived as pleasant. In our online survey the *Jungle* soundscape, which has the highest variety of sounds and reverb effects, was not rated higher than the *Forest* soundscape. We nevertheless assume that the preferences are user-dependent. However, the participants of the in-situ study stated to incorporate their visual impressions into the played soundscapes. For example, trees on the side of the road were woven into the acoustic forest and restaurants were integrated into an imagined walk down a beach promenade. To improve the overall experience, the selection of a soundscape could be based on visual impressions of the walked paths. Such "auditory anchors" may be stores and restaurants along the way or particular landmarks, which could be retrieved from map data.

From the moment at which step sounds were played, the environment was experienced as richer and more compelling. Augmented footstep sounds added a new quality to the experience. For instance, participants perceived the acoustic forest as more realistic when they heard the sound of their own footsteps on dry leaves; and P16 imagined to be back to a specific place of his childhood in the *Mountain* soundscape. In our interviews participants mentioned the word *focused* (on steps and breath) when augmented step sounds were played. Therefore, we wanted to find out whether the independent variables soundscapes and footstep sounds have effects on the users' step duration in the first study part. A Shapiro-Wilk test showed no normal distribution of the step duration with augmented footsteps in the soundscapes *Beach* and *Mountain*. Therefore, a Friedman test was performed that showed no significant differences between the three soundscapes with and without footstep augmentation regarding the step duration ($F(6{,}16) = 5.86, p = 0.32 > 0.05$). One reason for this outcome may be that the participants had to remain aware of their surroundings, e.g. to circumvent other pedestrians. However, this contradicts findings from related work in lab studies in which the step length and walking pace were both affected by different footstep sounds [Cor20; Tur13b]. In addition to external influences, the footstep sound volume was sometimes perceived to be too loud after initial attempts, although all sounds were normalized and related work found no significant difference in the preferred volume of soundscapes and footsteps [Tur13a]. Because a Shapiro-Wilk test confirmed a normal distribution of the volumes of soundscapes and footstep sounds in each study part, we performed a t-test for each pair of volumes. We found that each pair of soundscape volume and footstep volume was significantly different ($t(2{,}16) = 13.95, p < 0.001$). One-way ANOVAs were conducted to analyze the volume change of the soundscape and footstep sounds over the study tasks. The results revealed that once footstep sound volume is set, there is no further significant different change of the volume ($F(8{,}16) = 1.60, p = 0.14 > 0.05$). The same applies to the soundscape volume ($F(8{,}16) = 0.04, p = 0.99 > 0.05$). We therefore conclude, in contrast to [Tur13a], that the footstep volume should be adjusted below the soundscape volume. On average, the footstep volume ($\bar{vol} = 42.52\%; SD = 14.82$ %) was set 17.4 % below the soundscape volume ($\bar{vol} = 59.95\%; SD = 12.22$ %) for a pleasant experience. Although no other significant differences were found, a Mann-Kendall test found a significant decreasing trend for the footstep volume change over the study parts ($p = 0.006 < 0.05, z = 2.75, \tau = -0.16, s = -1332.0$) while no trend on the soundscape volume change could be identified ($p = 0.65 > 0.05, z = 0.45, \tau = -0.16, s = -0.03$). Figure 9.7 depicts the volume change of soundscapes and footsteps over the study.

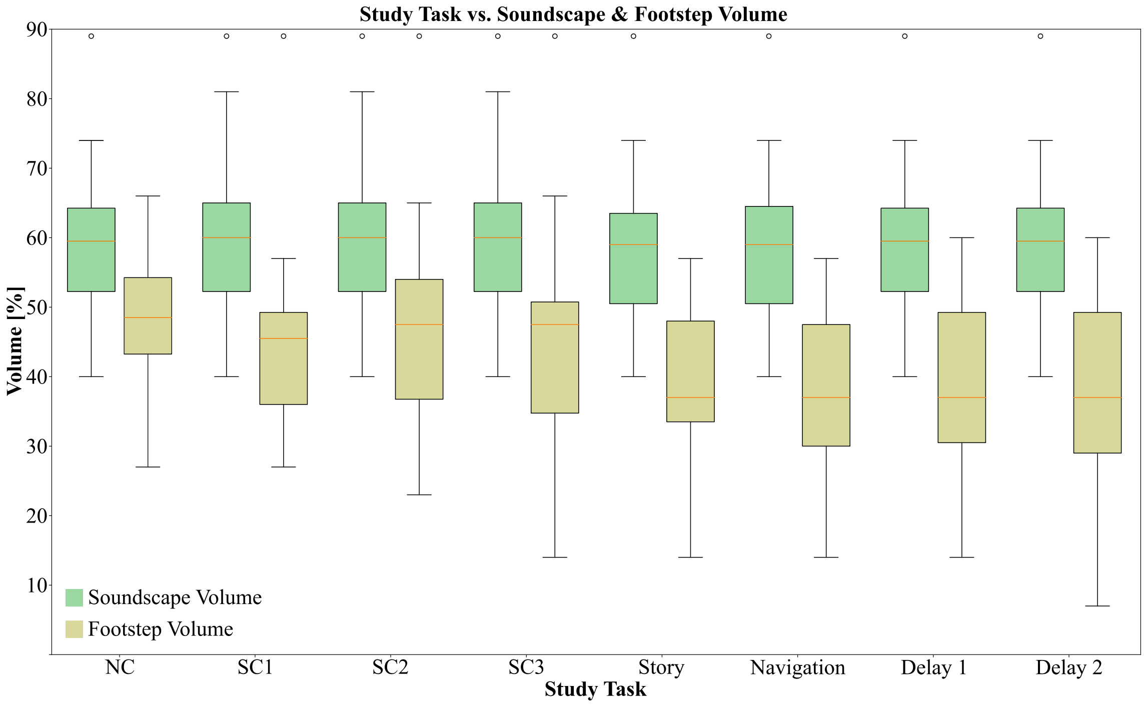

Figure 9.7: Volume adjustment over the course of the study: While both sounds were adjusted equally in the beginning, participants repeatedly lowered the footstep volume (in yellow) from an average of 60 % to 42.5 %.

Our participants reported that they perceived the footstep playback as more synchronous with their actual footsteps after having used the augmentation for a while. This is most likely not only due to a learning effect, but also depends on the volume of the step sounds. An asynchronous step feedback was also experienced as walking next to a person. According to P16, it took him 5 minutes to get used to the feedback and then he perceived the augmented footstep sounds as part of the soundscape. We conclude that a highly accurate footstep synchronization is not necessary in the context of AAR. In the closing questionnaire the synchronicity of the footstep sounds on a 5-point scale from 1 (fully asynchronous) to 5 (fully synchronous) has been evaluated (M=4.13, SD=0.93). Ten participants of the in-situ study rated the synchronicity as 4 (synchronous). Five participants stated full synchronicity (P3, P6, P7, P11, P16) and only P5 found the sounds to be fully asynchronous. Also for P5, a delay of the step sounds led to an increase in the step duration, according to a Mann-Kendall test ($p = 0.04 < 0.05, z = 1.97, \tau = 0.51, s = 23.0$), although he noticed no change in the rate of the footstep playback. Thus, we infer that delayed footstep playback may serve as an unobtrusive feedback method to slow pedestrians down. In contrast, a faster playback was perceived as stressful and pushing, which might be helpful in the context of running with the aim of holding a certain pace or to catch a bus during a navigation tasks. We also evaluated whether the participants were able to immerse themselves in the soundscapes on a 5-point scale from 1 (no immersion) to 5 (full acoustic immersion) (M=3.94, SD=0.66). Three participants totally agreed to have felt immersed into another world (P2, P6, P11), nine agreed and only four were neutral (P4, P8, P10, P13). Here we conclude that footstep augmentation contributes to immersing into audio-augmented realities. Commonly used questionnaires that measure immersion or presence [Wit98] were not applied, as augmented soundscapes fully preserve all visual stimuli of reality and thus several items on standard questionnaires do not apply [McG20]. Furthermore, the urban environment in the in-situ study presented an uncontrolled setting, leading to confounds when trying to measure immersion. We found that the participants combined external visual impressions and even wind sounds and their tactile and haptic perception into a unified experience. Therefore, we felt that interviews would be a more appropriate method to assess personal experiences of augmented soundscapes in public places [Bro14].

By telling the participants that they will hear an auditory story, they were easily able to make up story lines by themselves. Examples of stories they imagined were: a walk in Japan, being on a vacation, stranding on an lonely island, or even going back to their childhood. The auditory soundscape inspired the imagination of fictional stories or the evocation of memories, whereby the sounds of footsteps influence the imagination. Interestingly, none of our participants talked about a story in the third person. They were always referring to themselves as the main actor in the story, which suggests at least some degree of immersion into the played soundscapes. Mostly the participants imagined being surrounded by a natural environment without other people, which indicates that the participants mentally blocked out the crowded urban environment. Man-made warning sounds, like the bicycle bell and the ship horn, were perceived as surprising and could not be woven into their stories. Instead, those sounds brought the participants out of the story to focus more on the surrounding traffic. The woodpecker signal sound was a better fit for the soundscape, but it may lead to important warning signals being ignored. Hence, it seems more suitable for signaling a change of direction in a navigation task rather than as a warning sound. The binary fashion of the footstep feedback during the navigation task (standard sound when on the right track, wood sound when on the wrong track) was found to be sufficient to signal wrong walking directions. Although this simple navigation feedback was perceived as easy and intuitively understandable, it does not provide sufficiently strong cues to replace spatial audio navigation methods in terms of efficiency [Alb16; Hel20]. Thus, it should rather be used to deliver a stop signal, e.g. at a red traffic light, or to indicate when walking in the wrong direction. In terms of walking meditation, an audio augmented reality using footstep feedback can offer a new dimension of meditation. Individuals reported feeling more focused on their own steps and breath with footstep feedback. In addition, an AAR should not be considered as a replacement for walking meditation in urban soundscapes but as one potential element of a meditation method [Han11]. Further research is required to compare the effect of our approach on walking meditation and the effect of AAR without augmented footstep sounds.

In comparison to McGill et al. [McG20], our research confirms some of their results. Auditory displays can weave into the overall perceived environment. However, the visual influences and footstep sounds have an additional impact on how real an audio augmented reality is perceived. Embedded warning sounds can restore a certain perceived sense of safety with ANC headphones to make the user aware of their surroundings. Likewise, footstep sounds can be used to intuitively, instantly and partly unconsciously convey simple cues. This makes footstep sounds suitable to not overload an auditory display. However, our results show that these sounds require a short period of familiarization to be perceived as part of the overall reality. In addition, the footstep volume needs to be adjusted below the background soundscape volume to provide a more realistic experience, and faster step feedback should be avoided as it was perceived as stressful.

9.8 Evaluating Risk Factors in Urban Environments

The in-situ field study showed that pedestrians can react quickly to augmented acoustic signals. However, dangerous situations involving real risks of injury could not be evaluated. This section is intended to discuss what those risks are and how they could be evaluated in future research.

Figure 9.8 illustrates an example situation in which the proposed audio augmentation could support save use of noise canceling headphones in urban environments. In picture A the user walks along a street with ANC headphones on while a car approaches from out of sight. Similar situations can occur, for example, at parking exits or driveways. Illustration B depicts the corresponding audio augmented soundscape *Beach* where the user walks on sand and hears the ocean waves and seagulls. When the car is approaching, the user hears a ship horn from the direction of the car. The footstep sound changes to wooden planks to enforce the intention to stop and focus on the surroundings in the urban environment. Figure 9.8 C visualizes

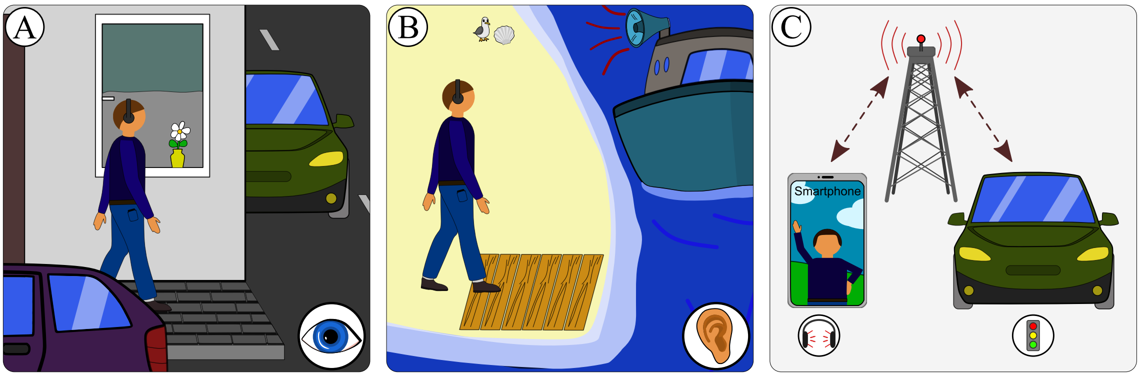

Figure 9.8: An illustration of an example danger situation. In A the user cant see the approaching. In B the user is made aware of an approaching danger by a played ship horn sound and the step sound of wooden planks. In C the visualization shows that the smartphone and the approaching car communicate via a cellular base station with each other.

how the car and smartphone of the user communicate via a cellular base station with each other. Both share their location and direction with a central computing system where potential collisions are calculated. In the field, this situation could not be easily evaluated because of involved risk of injuries. Apart from the risk of missing or ignoring important signals, the computational effort and transmission delays can cause signals to reach the user too late.

Virtual realities make evaluations involving dangerous situations possible. To exemplify, Chang et al. [Cha17a] evaluated whether eyes on cars support communication with pedestrians. They found that pedestrians felt safer when eyes indicate the intention of approaching cars. However, the proposed scenario in Figure 9.8 A limits the capabilities of their approach. When obstacles prevent direct line of sight, the auditory channel can become useful for communication. Our implemented VR environment is made with Unity 3D [Tec22] and the freely available virtual environment asset *Otaku City* [ZEN]. We animated driving cars and walking pedestrians from *Mixamo* [Inc] to create a lively urban environment. For moving through the environment, we implemented a test application for Android based on the functions used for *EnvironZen*. Detected steps are transmitted via WiFi to the Oculus Quest 2 VR headset. The user physically walks on a sport while virtually walking in the current virtual viewing direction. To trigger sounds, the VR headset sends WiFi commands to the smartphone placed in the trouser pocket. The commands can change the soundscape, footstep sounds and trigger directional warning sounds as well as navigation sounds.

Figure 9.9 shows images of the implemented urban environment and situations that can be used to evaluate whether users trust an audio-augmented reality. Beginning from the starting point in A, the user walks along a lively street and hears the

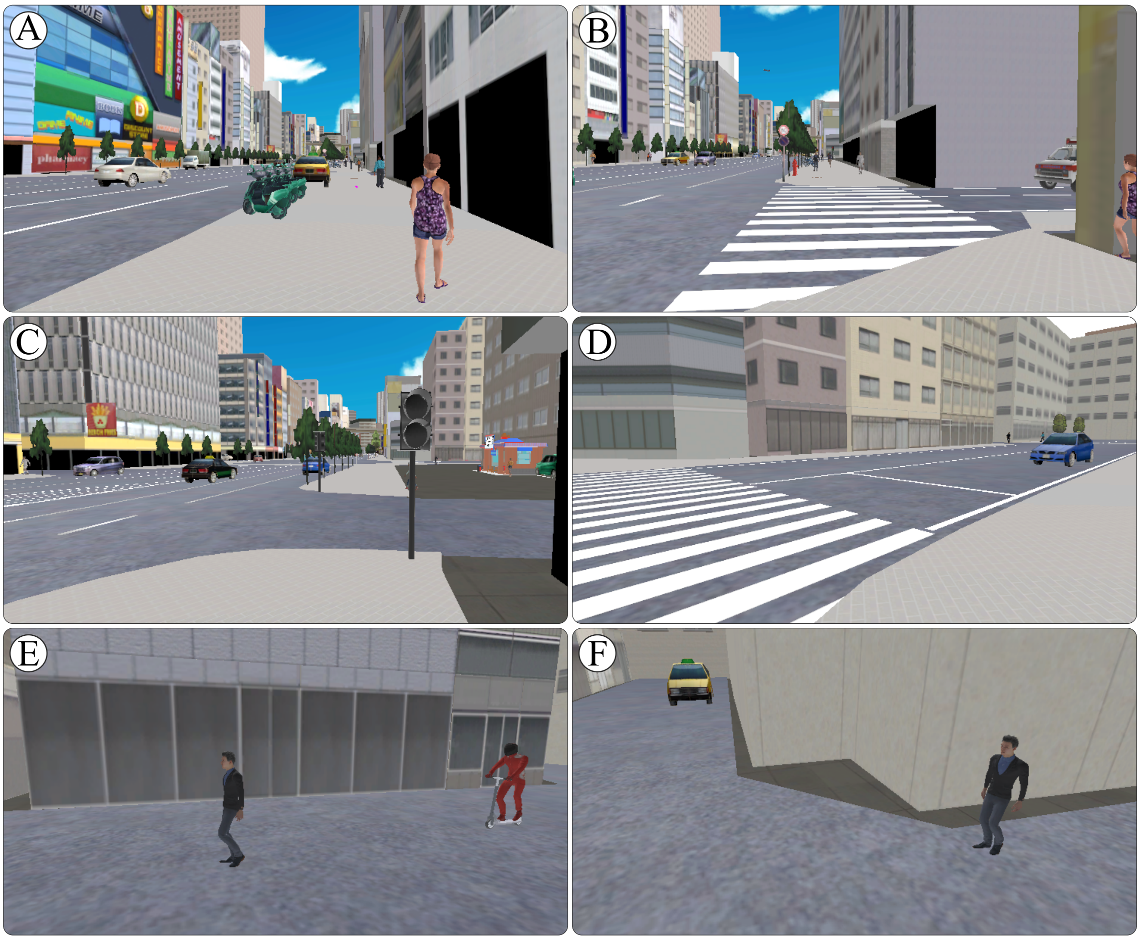

Figure 9.9: Images of the implemented urban VR environment. Picture A shows the starting point where the user walks along a walkway. In B the first dangerous situation is shown with an approaching ambulance out of sight on a crosswalk. In C the traffic light is out of order. D implements a comparable scenario to [Cha17a]. In E a scooter driver is approaching from behind and F presents a driveway scenario.

augmented footstep sounds and soundscape *Forest*. Thus, the user can get familiar with the implemented system. When coming to the crosswalk in B, the step sound changes to wooden planks and the warning sound *Woodpecker* is triggered to alert the user of the approaching ambulance. The intention is to evaluate how users behave when the AAR sounds change. From the visual perspective the user sees a crossroad giving him the intention that cars stop when walking. However, the situation changes with an approaching ambulance out of sight. It could possibly happen in the field that one is walking with ANC headphones but overhears the siren due to a loud volume level and/or a lack of awareness of the environment. With our proposed system it is now of interest whether users stop or keep on walking. In C the traffic light is out of order but the step sounds do not change. Also here, it is of interest do users continue walking and trust the AAR system or do they stop and look around to make sure they can safely cross the street. Situation D implements a comparable scenario to [Cha17a] but with our proposed technology. Here, cars are approaching continuously without changing footstep sounds or warning sounds. When the user enters the cross road cars stop immediately. It is of interest whether users wait for a gap between driving cars to cross the street or will they simply continue walking. In image E an e-scooter driver is approaching towards the user from the back which is signaled by a directional warning sound. This scenario is comparable to cases from the in-situ field study. It is expected that users look behind when they hear a warning sound. The last scene in picture F implements a comparable situation to Figure 9.8A. A car is approaching from a parking exit or driveway and warning sounds as well as an altered step sound hint the user to stop walking.

9.9 Conclusion

This chapter presented the results of an online survey and a subsequent in-situ field study on augmenting footstep sounds in audio-augmented realities. The field study took place in an urban environment. The soundscapes (*Forest*, *Beach*, and *Mountain*) and footstep sounds (*Leaves*, *Sand*, and *Gravel*) that received the highest ratings in the online survey were investigated in detail regarding their effects on the experience of pedestrians on a sidewalk. Audio was played with noise-canceling headphones and footsteps were detected with the accelerometer of a mobile phone worn in the trouser pocket.

We found that augmented footstep sounds contribute to the experience of presence in acoustic environments. Augmented footstep sounds create a link between acting in the physical environment – performing the footsteps, experiencing the haptic feedback, and visually perceiving one's movement in the urban space – and in the virtual environment – hearing the virtual footstep sounds in the context of the soundscape. So users act simultaneously in the physical and the virtual world. It seems that the sounds of the preferred soundscapes and of the augmented footsteps are sufficiently indeterminate that users are able to integrate them with the other stimuli to a coherent whole. Users seem to be able to concurrently allocate attention to the auditory stimuli from the soundscape, the virtual sounds created by their footsteps, and the visual stimuli from the physical environment. Here, the imaginations were very different from person to person. When soundscapes are sequentially played, listeners are able to connect them to form stories in urban environments, in which they themselves act as the main characters. Soundscapes can also evoke memories like vacations and these memories or imaginations may even be the main factor determining the level of immersion in quiet, acoustic compositions.

Adjusting the volume of footstep sounds was found to be necessary for creating pleasurable, individual experiences. Slightly delayed footstep playback showed a significant impact on the step duration, making it possible to slow people down, while faster footstep playback showed no impact on walking speed but led to users feeling stressed and rushed. Slower footstep sounds were perceived as relaxing and calming. Using footstep sounds for navigation purposes can be intuitively understood, but is limited in effectively conveying a sense of direction. However, changing the footstep sound can be used to indicate the wrong direction, red traffic lights, or the right way with only two choices. Augmenting further sounds with spatial audio in soundscapes, like approaching vehicles, can be easily understood but should be introduced beforehand for not being misunderstood. One advantage of audio-only augmentation for public spaces is that visual stimuli are fully preserved, so that possible harms only have to be augmented if out of sight of the user.

The underlying algorithm for real-time footstep playback was found to be mostly accurate in the study, but could be adapted to each user to improve the reliability of the acoustic immersion. One advantage of the approach is that it can be implemented and used on any smartphone without additional equipment. *EnvironZen* is freely available at GitHub [1] and open-sourced for future research. As people need to continue to pay attention to events in the physical environment, safety is an important issue to be further explored in the future. Thus, we motivated several dangerous situations that could be evaluated in virtual environments. We created an urban virtual test environment based on freely available components and implemented several situations in which our proposed system could support safe guidance of pedestrians with ANC headphones. In the future, this implementation could serve to evaluate trust of our proposed AAR system.

1 https://github.com/M-Schrapel/EnvironZen

CHAPTER 10

Conclusion

Mark Weiser's vision of ubiquitous computing has contributed to the integration of computers into the physical world [Wei91]. Surface computing has been more focused on embedded hardware into everyday objects for a long time. In the chapters of this thesis, we have examined surface computing from a different perspective. Instead of placing computational systems and knowledge in the world, we used already existing knowledge for computational systems and human-computer interactions. Mobile technology served to measure, use and augment surfaces in the environment. This, perhaps at first glance, small difference can result in better scalability of ubiquitous computing. We will now conclude and summarize our findings by discussing our results in relation to our initial formulated hypotheses from Chapter 1.

H1 Surface computing can be subdivided into active and passive surface computing.

H2 Active surface interactions are convertible into passive surface interactions and vice versa.

Our first two hypotheses served to verify that surface computing can be divided into active and passive surface computing. We could find evidence in related work that surface computing can be divided into active and passive surface approaches. We also found approaches that had similar goals but could be divided into active and passive surface computing. An example was electrovibration (active) [Bau10] and inverse electrovibration (passive) [Bau12]. In addition, this dissertation has also provided several contrast examples to active surfaces interactions. For instance, in Chapter 9 we presented an approach that augments footstep sounds in urban environment using ANC headphones. In related work we find an approach where roads are equipped with sensors and loudspeakers to achieve similar goals [Ren21]. Both approaches of surface computing have advantages as well as disadvantages which are also mentioned in Chapter 2. While equipment on the running route results in the user not having to carry any equipment, our approach requires the user to wear mobile technology but is more resource-efficient and universally applicable. However, the context determines whether passive surfaces are to used for an interaction and which task is to be carried out. Nevertheless, active and passive surfaces should not be considered as competing approaches. The decision is up to a desired user group and application scenario. For a runner in the forest, there is no obvious reason to wear a headset for hearing the soundscape and footsteps of a forest. In addition, data of the user and the environment must be collected to augment a passive surface. This can lead to privacy concerns, which have been analyzed over the contributing studies of this thesis. It turns out that cameras and location data result in the greatest privacy concerns, followed by audio data and motion data with the least concerns. This means that our approach is limited to the extent that a user is willing to share the collected data. However, data are indispensable for an integration of passive surfaces in human-computer interactions. We have further analyzed how measured data of an ordinary surface contribute to recognize an interaction and formulated related hypothesis.

H3 Passive surfaces contain information about the type of interaction that is performed on it.

H4 Passive surfaces contain information about the user that performs an interaction.

H5 Vibrations combined with scratching sounds produced by friction on a surface contain complementary features that contribute to identify an interaction.

H6 Interactions with passive surfaces emit information about the type of surface on which the interaction is performed.

We have shown by the application scenario of handwriting recognition in Chapter 3 that scratching sounds and vibrations when moving on a surface can support to recognize an interaction. A particular use case that proofed the validity of H3 and H5 is the recognition of different digits like “6” and “0”. By integrating scratching sounds into the classifier both digits could be recognized at higher accuracy. Both digits have similar writing trajectories resulting in similar related motion data. However, scratching sounds produced by the surface served to better distinguish both digits. This observation could be confirmed in chapter 4 through the identification of users by single handwritten symbols. Our classifiers were considerably more accurate when audio and motion data were combined as input for an SVM. The pressure on a surface, related scratching sounds and the acoustic feedback channel over the bones turned out to be individual among a small user group of 30 persons which resulted in high recognition rates of the writers even though we collected a small data set. Furthermore, we were able to show in Chapter 5 that the skin on the back of the hand can serve as a suitable ubiquitous input surface. In a usability experiment, it was confirmed that two-dimensional interactions directly on the skin are perceived as more intuitive than mid-air interactions. In the related work similar results can be found for virtual keyboards [Dud19]. Chapters 6 and 7 contributed to evaluate H6. In Chapter 7, we could identify surfaces by texture and color features. In Chapter 6 we combined audio and motion data of a sensor kit in two application scenarios. For the cycling scenario, we could show that combined sensor data reach higher recognition rates than single sensor approaches. This shows that H6 can be confirmed by recognizing different road conditions. Furthermore, it was shown that route preferences are weather-dependent. However, the use case of skiing and measuring different slope conditions turned out to be challenging due to the high dynamic changes on the slopes. A ski slope can have highly dynamic, changing conditions that are more challenging to distinguish than in a cycling scenario. We assume that a larger amount of data and a more accurate labeling procedure are necessary to achieve meaningful results. However, as with cycling, preferences for surface conditions were found among skiers. This lead to the following hypotheses.

H7 Passive surfaces and their properties in the wild can support users in their everyday tasks.

H8 Passive surfaces are able to sense the context of using technology.

We have examined H7 in various application scenarios. At first we focused at digital pens using audio and motion data. It was found that surface related features are sufficient to achieve accuracies of 98 % on handwritten digits. High accuracies on recognized numbers are necessary because digits, unlike words, cannot be corrected afterwards by auto-correction. Furthermore, we could show in Chapter 4 that these data of handwritings are also able to recognize the writer. We have motivated that forged signatures can be identified through scratching sounds and the style of holding the pen. In Chapter 5 we showed that magnetic tracking can enhance the interaction space of small wrist-worn displays. Thereby, the back of the hand was used as an interactive space similar to a touch pad. In Chapter 6 we found that road and slope conditions influence preferred route choices. In Chapter 7, we discussed that surfaces contribute to sensing the context of smartphone usage. We envisioned an application that recognizes the surface a smartphone is placed on. This app served to verify H8 by determining whether a person is working at an office desk or taking a break at an coffee table. We motivated that sensed surface information can be used to manage notifications in order to work more efficiently. Chapter 8 showed that the information from book spines can be recognized by smartphone cameras to find a desired book on the shelves more efficiently. We envisioned that visual highlighting of the book spine could possibly facilitate book search for online sharing communities. In addition to visual augmentation, we have explored acoustical augmented surface information to alter (passive) surface perception in the wild. We also considered how these altered surfaces can be used for interactions and formulated two hypotheses.

H9 Acoustically augmented surfaces influence the perception of an overall perceived reality.

H10 Acoustically augmented surfaces can communicate special cues and change the behavior of the user.

In Chapter 9 we investigated several effects of acoustical surface augmentation in urban environments. By interviewing our participants, we found that virtual soundscapes were influenced by the perception of environmental stimuli. For example, restaurants were woven into individual perceptions leading to the imagination of a beach promenade when the soundscape beach was played. The augmentation of footstep sounds in the forest soundscape made study participants feel more deeper in the forest and participants reported to feel more focused (on steps and breath) with footstep feedback. Therefore, and in relation to H9, we deduce that augmented footstep sounds influence the perception of the perceived overall reality. Furthermore, we found that altered footstep feedback can influence walking speed. A faster footstep playback showed no effect and was partly perceived as stressful. In contrast a slower footstep playback showed effects on the walking speed. Our participants were slowed down although they have partially noticed no differences in the footstep feedback. In addition, changing footstep sounds were found to be sufficient for conveying simple navigational cues such as stop signals or indications that a person is walking in the wrong direction. An augmentation of footstep sounds can thus contribute to influencing behavior in urban environments, which confirms H10.

10.1 Limitations

This thesis has shown several limitations of passive surface computing. First of all, mobile sensors that permanently measure the environment may raise privacy concerns. We surveyed our participants in the different studies about their privacy concerns regarding our used sensors. We found a tendency among all studies for motion data to cause the least concern, followed by audio data and video data, which raised the most concerns. In Chapter 6 we also surveyed location data (GPS), which were ranked to raise less concerns than videos. Sensors are indispensable to integrate common surfaces into an interaction. Since the environment is measured, not only one's own privacy is affected, but also that of the people in the immediate vicinity.

We have measured audio data via contact microphones in order to attenuate noise from the environment. However, it cannot be completely excluded that ambient noises or spoken words influence our audio measurements. This can result in a lower recognition of an interaction when using audio data [Mit19]. For this reason, we have limited our results regarding scratching sounds from digital pens to quiet environments. In addition, audio signals require considerably higher sample rates than motion signals. Therefore, larger amounts of data are usually required before a classifier can reliably detect an interaction or surface. In the case of writer identification, we were able to show that transferring audio data into the frequency domain can establish reliable results with small amounts of data. However, in the use case of skiing, we could not achieve sufficient recognition rates to distinguish snow types. The limited amount of data as well as the highly dynamic changes on the slope were crucial for our results.

The use of cameras to detect and use surfaces for interactions also shows limitations. Besides privacy concerns, a clear visibility of the surface is necessary. In the case of book cases, it was shown that the recognition of books on the shelves is lower in outdoor lighting conditions. Incident light changes the brightness of the book spines and thus the measured color. We used thresholds to pre-select books by color features and then highlighted a potential target book by text features. If there is no text on the book spine, false detections are more likely to occur. In Chapter 7 we measured the surface on which a smartphone is placed by flashlights and rear cameras. For similar colored surfaces and especially for black surfaces, we have noticed larger confusions. Therefore, it can be assumed that the approach must be limited to a heterogeneous color distribution, as in the case of our book spine algorithm.

Lastly, we examined audio augmented surfaces. Our approach could not produce instant footstep feedback. In addition, the rolling of the foot on the surface could not be reproduced by the played footstep sounds. The in-situ study showed that after a familiarization phase, minor delays are no longer noticed and the step sounds are perceived as part of the overall reality. However, one participant of the in-situ study perceived the playback to be asynchronous with little improvements over the study. In addition, step sounds were found to be an appropriate feedback method for navigation tasks, but the feedback was limited to conveying simple cues such as stop signals.

10.2 Future Work

This thesis investigated the integration of ordinary surfaces into human-computer interactions. Although surfaces are shown to have a wide range of potential applications, the effects have been measured over relatively short time periods. This suggests that further long-term studies are required with our presented prototypes to draw in-depth conclusions. In addition, it is necessary to further investigate the robustness of scratching sounds in relation to ambient noise.

We assume that augmented reality will continue to gain in importance. New display technologies could enable more realistic image superimpositions in the field of view [Ito21; Xio21]. As a result visual approaches of passive surface interactions could gain more research interest over the next years. To enable interactions, virtual touch detection algorithms have to become more efficient. Likewise, digital pens are an ongoing research topic. In addition to applications in teaching undergraduate students [Gmb22], an increasing amount of proposed pen applications for virtual reality interactions indicates the need for further research on pen interactions [Dre20; Ges20; Rom21]. Moreover, the trend of sustainable multi-modal mobility and IoT sensor networks could lead to a variety of vehicles in urban areas that measure road conditions to improve maps.

10.3 Closing Remarks

Ordinary surfaces surround us at any time and everywhere. They contain information available directly at our fingertips. The goal of this thesis was to use and augment their different properties for human-computer interactions. We introduced that surface computing can be divided into active and passive surface computing. While active surfaces embed hardware in or near the interaction area, passive surface computing utilizes mobile technology to enable interactions with computers anywhere and on any surface. Without instrumenting the environment, our presented approach is a step towards making our everyday lives easier while technology fades from the forefront. Starting from measuring, using and augmenting surface related features up to the identification of an interaction or the users themselves, we have presented a wide range of application scenarios. However, users must be willing to use mobile technology and share collected data. In addition, the context may be a determining factor in whether and what interaction with computers is carried out.

Bibliography

[Aar09] Aarts, Emile and Boris De Ruyter: 'New research perspectives on ambient intelligence'. *Journal of Ambient Intelligence and Smart Environments* (2009), vol. 1(1): pp. 5–14 (cit. on p. 16).

[Aar03] Aarts, Emile and Stefano Marzano: *The new everyday: Views on ambient intelligence*. 010 Publishers, 2003 (cit. on p. 16).

[Abd20] Abdelnasser, H., K. A. Harras, and M. Youssef: 'MagStroke: A Magnetic Based Virtual Keyboard for Off-the-Shelf Smart Devices'. *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. 2020: pp. 1–9 (cit. on p. 72).

[Abe16] Abe, Tetsuya, Buntarou Shizuki, and Jiro Tanaka: 'Input Techniques to the Surface Around a Smartphone Using a Magnet Attached on a Stylus'. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '16. San Jose, California, USA: ACM, 2016: pp. 2395–2402 (cit. on pp. 72, 73).

[Abo99] Abowd, Gregory D, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles: 'Towards a better understanding of context and context-awareness'. *International symposium on handheld and ubiquitous computing*. Springer. 1999: pp. 304–307 (cit. on p. 17).

[Abo00] Abowd, Gregory D and Elizabeth D Mynatt: 'Charting past, present, and future research in ubiquitous computing'. *ACM Transactions on Computer-Human Interaction (TOCHI)* (2000), vol. 7(1): pp. 29–58 (cit. on p. 16).

[Ach21] Acharya, Sraddhanjali and Abdul Serwadda: 'On Finger Stretching and Bending Dynamics as a Biometric Modality'. *2021 International Conference on Smart Applications, Communications and Networking (SmartNets)*. IEEE. 2021: pp. 1–8 (cit. on p. 56).

[Ack07] Ackery, Alun D., Brent E. Hagel, Christine Provvidenza, and Charles Tator: 'An international review of head and spinal cord injuries in alpine skiing and snowboarding'. *Injury Prevention* (2007), vol. 13: pp. 368–375 (cit. on p. 94).

[Ade01] Adelson, Edward H: 'On seeing stuff: the perception of materials by humans and machines'. *Human vision and electronic imaging VI*. Vol. 4299. SPIE. 2001: pp. 1–12 (cit. on p. 11).

[Aer95] Aeronautics, National and Space Administration: 'NASA-STD-3000, Man-System Integration Standards, Volume 1, Section3 : Anthropometric Dimensional Data'. (1995), vol. 1 (cit. on pp. 74, 75, 79).

[al15] al., Martin Abadi et: *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015 (cit. on p. 135).

[Ala20] Alajrami, Eman, Belal AM Ashqar, Bassem S Abu-Nasser, Ahmed J Khalil, Musleh M Musleh, Alaa M Barhoom, and Samy S Abu-Naser: 'Handwritten signature verification using deep learning'. *International Journal of Academic Multidisciplinary Research (IJAMR)* (2020), vol. 3(12) (cit. on p. 55).

[Alb16] Albrecht, Robert, Riitta Väänänen, and Tapio Lokki: 'Guided by music: pedestrian and cyclist navigation with route and beacon guidance'. *Personal and Ubiquitous Computing* (Feb. 2016), vol. 20 (cit. on pp. 155, 175).

[Ale17] Alessandroni, Giacomo, Alberto Carini, Emanuele Lattanzi, Valerio Freschi, and Alessandro Bogliolo: 'A study on the influence of speed on road roughness sensing: The SmartRoadSense case'. *Sensors* (2017), vol. 17(2): p. 305 (cit. on p. 93).

[Alo14] Alonso, J., J.M. Lopez, I. Pavon, M. Recuero, C. Asensio, G. Arcas, and A. Bravo: 'On-board wet road surface identification using tyre/road noise and Support Vector Machines'. *Applied Acoustics* (2014), vol. 76: pp. 407–415 (cit. on p. 94).

[Alp92] Alpaydin, E.: 'Multiple neural networks and weighted voting'. *Proceedings., 11th IAPR International Conference on Pattern Recognition. Vol.II. Conference B: Pattern Recognition Methodology and Systems*. 1992: pp. 29–32 (cit. on p. 42).

[Alv10] Alvarsson, Jesper J, Stefan Wiens, and Mats E Nilsson: 'Stress recovery during exposure to nature sound and environmental noise'. *International journal of environmental research and public health* (2010), vol. 7(3): pp. 1036–1046 (cit. on pp. 155, 156).

[Alz12] Alzantot, Moustafa and Moustafa Youssef: 'UPTIME: Ubiquitous pedestrian tracking using mobile phones'. *2012 IEEE Wireless Communications and Networking Conference (WCNC)*. 2012: pp. 3204–3209 (cit. on p. 159).

[And09] Anderson, Barton L and Juno Kim: 'Image statistics do not explain the perception of gloss and lightness'. *Journal of vision* (2009), vol. 9(11): pp. 10–10 (cit. on p. 11).

[Ann18] Anne Finger and Lena Greinke and Maximilian Schrapel: 'Fußverkehr als Beitrag zur Gesunden Stadt'. *PLANERIN 5/2018: Gesundheitsversorgung - Gesund leben in Stadt, Land und Quartier* (2018), vol. 5 (cit. on p. 268).

[AB22] AB, Anoto: *Anoto. Digital pen and paper technology.* http://www.anoto.com. 2022 (cit. on pp. 21, 32, 56, 66).

[Ant93] Antonakos, Cathy L: 'Environmental and travel preferences of cyclists'. PhD thesis. University of Michigan, 1993 (cit. on p. 92).

[AP93] AP, Leppävuori, Karras M, Rusko H, and Viitasaio JT.: 'A New Method of Measuring 3-D Ground Reaction Forces under the Ski during Skiing on Snow'. *Journal of applied biomechanics* (Jan. 1993), vol. 4(9): pp. 315–328 (cit. on p. 95).

[App18] Apple: *Apple Developer Documentation.* https://developer.apple.com/documentation/. Accessed: 2021-02-02. 2018 (cit. on pp. 118, 127).

[Ara97] Arai, Toshifumi, Dietmar Aust, and Scott E. Hudson: 'PaperLink: A Technique for Hyperlinking from Real Paper to Electronic Content'. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems.* CHI '97. Atlanta, Georgia, USA: ACM, 1997: pp. 327–334 (cit. on p. 33).

[Ark99] Ark, Wendy S. and Ted Selker: 'A look at human interaction with pervasive computers'. *IBM systems journal* (1999), vol. 38(4): pp. 504–507 (cit. on p. 16).

[Asa13] Asai, Hiroki and Hayato Yamana: 'Detecting Student Frustration Based on Handwriting Behavior'. *Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software and Technology.* UIST '13 Adjunct. St. Andrews, Scotland, United Kingdom: ACM, 2013: pp. 77–78 (cit. on pp. 40, 58).

[Ash11] Ashbrook, Daniel, Patrick Baudisch, and Sean White: 'Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring'. May 2011: pp. 2043–2046 (cit. on p. 72).

[Aug10] Augsten, Thomas, Konstantin Kaefer, René Meusel, Caroline Fetzer, Dorian Kanitz, Thomas Stoff, Torsten Becker, Christian Holz, and Patrick Baudisch: 'Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input'. *Proceedings of the 23nd annual ACM symposium on User interface software and technology.* 2010: pp. 209–218 (cit. on p. 20).

[Aut] Autonomio: *Talos: Hyperparameter Optimization for Keras.* `http://github.com/autonomio/talos`. Accessed: 2020-09-01 (cit. on p. 100).

[Bai11] Bailador, Gonzalo, Carmen Sanchez-Avila, Javier Guerra-Casanova, and Alberto de Santos Sierra: 'Analysis of pattern recognition techniques for in-air signature biometrics'. *Pattern Recognition* (2011), vol. 44(10-11): pp. 2468–2478 (cit. on p. 55).

[Bal06] Ballagas, R., J. Borchers, M. Rohs, and J.G. Sheridan: 'The smart phone: a ubiquitous input device'. *IEEE Pervasive Computing* (2006), vol. 5(1): pp. 70–77 (cit. on pp. 1, 24, 94, 116, 158).

[Ban20a] Bandukda, Maryam and Catherine Holloway: 'Audio AR to Support Nature Connectedness in People with Visual Disabilities'. UbiComp-ISWC '20. Virtual Event, Mexico: Association for Computing Machinery, 2020: pp. 204–207 (cit. on p. 155).

[Ban20b] Bandukda, Maryam, Catherine Holloway, Aneesha Singh, and Nadia Berthouze: 'PLACES: A Framework for Supporting Blind and Partially Sighted People in Outdoor Leisure Activities'. *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility.* ASSETS '20. Virtual Event, Greece: Association for Computing Machinery, 2020 (cit. on pp. 155, 156).

[Ban03] Bang, Won-Chul, Wook Chang, Kyeong-Ho Kang, Eun-Seok Choi, Alexey Potanin, and Dong-Yoon Kim: 'Self-contained Spatial Input Device for Wearable Computers'. *Proceedings of the 7th IEEE International Symposium on Wearable Computers.* ISWC '03. Washington, DC, USA: IEEE Computer Society, 2003: pp. 26– (cit. on p. 33).

[Bar17] Barclay, Leah: 'Augmenting Urban Space with environmental Soundscapes and mobile technologies'. *Soundscape. Sounds Emergent: Diverse Ecologies Part II* (2017), vol. 16: pp. 20–34 (cit. on p. 156).

[Bar12a] Bardhi, Fleura and Giana Eckhardt: 'Access-Based Consumption: The Case of Car Sharing'. *Journal of Consumer Research* (Dec. 2012), vol. 39: pp. 881–898 (cit. on p. 131).

[Bar12b] Baroni, Adone, Elisabetta Buommino, Vincenza De Gregorio, Eleonora Ruocco, Vincenzo Ruocco, and Ronni Wolf: 'Structure and function of the epidermis related to barrier properties'. *Clinics in dermatology* (2012), vol. 30(3): pp. 257–262 (cit. on p. 14).

[Bas20] Basdogan, Cagatay, Frédéric Giraud, Vincent Levesque, and Seungmoon Choi: 'A review of surface haptics: Enabling tactile effects on touch surfaces'. *IEEE transactions on haptics* (2020), vol. 13(3): pp. 450–470 (cit. on pp. 21, 22).

[Bas12] Bashir, Muzaffar and Florian Kempf: 'Advanced biometric pen system for recording and analyzing handwriting'. *Journal of Signal Processing Systems* (2012), vol. 68(1): pp. 75–81 (cit. on p. 67).

[Bas08] Bashir, Muzaffar and Jürgen Kempf: 'Reduced Dynamic Time Warping for Handwriting Recognition Based on Multi dimensional Time Series of a Novel Pen Device'. (Jan. 2008), vol. 3 (cit. on pp. 33, 34, 56).

[Bas09] Bashir, Muzaffar, Jürgen Kempf, G Schickhuber, and Georg Scharfenberg: 'Online person authentication using dynamic signature on a novel tactile and pressure sensitive pad'. *17th Telecommunication forum, TELFOR Belgrade, Serbia*. 2009 (cit. on p. 56).

[Bas11] Bashir, Muzaffar, Georg Scharfenberg, and Jürgen Kempf: 'Person Authentication by Handwriting in air using a Biometric Smart Pen Device.' Vol. 191. Jan. 2011: pp. 219–226 (cit. on p. 56).

[Bat20] Batool, Faiza Eba, Muhammad Attique, Muhammad Sharif, Kashif Javed, Muhammad Nazir, Aaqif Afzaal Abbasi, Zeshan Iqbal, and Naveed Riaz: 'Offline signature verification system: a novel technique of fusion of GLCM and geometric features using SVM'. *Multimedia Tools and Applications* (2020), vol.: pp. 1–20 (cit. on p. 55).

[Bau12] Bau, Olivier and Ivan Poupyrev: 'REVEL: tactile feedback technology for augmented reality'. *ACM Transactions on Graphics (TOG)* (2012), vol. 31(4): pp. 1–11 (cit. on pp. 27, 181).

[Bau10] Bau, Olivier, Ivan Poupyrev, Ali Israr, and Chris Harrison: 'TeslaTouch: electrovibration for touch surfaces'. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010: pp. 283–292 (cit. on pp. 3, 22, 23, 27, 181).

[Bau09] Baudisch, Patrick and Gerry Chu: 'Back-of-Device Interaction Allows Creating Very Small Touch Devices'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Boston, MA, USA: Association for Computing Machinery, 2009: pp. 1923–1932 (cit. on p. 72).

[Bec19] Becker, Vincent, Linus Fessler, and Gábor Sörös: 'GestEar: Combining Audio and Motion Sensing for Gesture Recognition on Smartwatches'. *Proceedings of the 23rd International Symposium on Wearable Computers*. ISWC '19. London, United Kingdom: Association for Computing Machinery, 2019: pp. 10–19 (cit. on p. 56).

[Bed95] Bederson, Benjamin B: 'Audio augmented reality: a prototype automated tour guide'. *Conference companion on Human factors in computing systems*. 1995: pp. 210–211 (cit. on p. 155).

[Bel19] Bello, Juan P., Claudio Silva, Oded Nov, R. Luke Dubois, Anish Arora, Justin Salamon, Charles Mydlarz, and Harish Doraiswamy: 'SONYC: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution'. *Commun. ACM* (Jan. 2019), vol. 62(2): pp. 68–77 (cit. on p. 155).

[Ben14] Benfield, Jacob A, B Derrick Taff, Peter Newman, and Joshua Smyth: 'Natural sound facilitates mood recovery'. *Ecopsychology* (2014), vol. 6(3): pp. 183–188 (cit. on pp. 155, 156).

[Ben08] Benko, Hrvoje, Andrew D Wilson, and Ravin Balakrishnan: 'Sphere: multi-touch interactions on a spherical display'. *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 2008: pp. 77–86 (cit. on p. 18).

[Ber99] Berglund, Birgitta, Thomas Lindvall, Dietrich H Schwela, World Health Organization, et al.: 'Guidelines for community noise'. (1999), vol. (cit. on p. 154).

[Ber03] Bergstrø"m, A and R Magnusson: 'Potential of transferring car trips to bicycle during winter'. *Transportation Research Part A: Policy and Practice* (2003), vol. 37(8): pp. 649–666 (cit. on p. 94).

[Ber19] Bergström, Joanna and Kasper Hornbæk: 'Human–Computer Interaction on the Skin'. *ACM Computing Surveys (CSUR)* (2019), vol. 52(4): pp. 1–14 (cit. on p. 25).

[Bet13] Betsworth, Liam, Nitendra Rajput, Saurabh Srivastava, and Matt Jones: 'Audvert: Using spatial audio to gain a sense of place'. *IFIP Conference on Human-Computer Interaction*. Springer. 2013: pp. 455–462 (cit. on p. 155).

[Bia95] Biasca, N, H Battaglia, HP Simmen, P Disler, and O Trentz: 'An overview of snow-boarding injuries'. *Der Unfallchirurg* (Jan. 1995), vol. 98(1): pp. 33–39 (cit. on pp. 92, 95).

[Bib20] Bibi, Kiran, Saeeda Naz, and Arshia Rehman: 'Biometric signature authentication using machine learning techniques: Current trends, challenges and opportunities'. *Multimedia Tools and Applications* (2020), vol. 79(1): pp. 289–340 (cit. on p. 55).

[Bic08] Bicchi, Antonio, Martin Buss, Marc O Ernst, and Angelika Peer: 'The sense of touch and its rendering'. *Springer Tracts in Advanced Robotics* (2008), vol. 45 (cit. on p. 9).

[Bij10] Bijsterveld, Karin: 'Acoustic cocooning: How the car became a place to unwind'. *The Senses and Society* (2010), vol. 5(2): pp. 189–211 (cit. on p. 155).

[Bíl15] Bíl, Michal, Richard Andrášik, and Jan Kubecek: 'How comfortable are your cycling tracks? A new method for objective bicycle vibration measurement'. en. *Transportation Research Part C: Emerging Technologies* (2015), vol. 56: pp. 415–425 (cit. on p. 93).

[Bil15] Billinghurst, Mark, Adrian Clark, Gun Lee, et al.: 'A survey of augmented reality'. *Foundations and Trends® in Human–Computer Interaction* (2015), vol. 8(2-3): pp. 73–272 (cit. on p. 29).

[Blu] Bluetooth SIG, Inc.: *Bluetooth 5.1 Core Specification*. https://vtsociety.org/wp-content/uploads/2019/07/Core_v5.1.pdf. Accessed: 2022-02-23 (cit. on p. 128).

[Bol18] Boletsis, Costas and Dimitra Chasanidou: 'Smart tourism in cities: Exploring urban destinations with audio augmented reality'. *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. 2018: pp. 515–521 (cit. on p. 155).

[Bor14] Boren, B., M. Musick, J. Grossman, and Agnieszka Roginska: 'I Hear NY4D: Hybrid Acoustic and Augmented Auditory Display for Urban Soundscapes'. English (US). *International Conference on Auditory Display*. June 2014 (cit. on pp. 155, 156).

[Bos12] Bosco, Ruggero, Jeroen Van Den Beucken, Sander Leeuwenburgh, and John Jansen: 'Surface engineering for bone implants: a trend from passive to active surfaces'. *Coatings* (2012), vol. 2(3): pp. 95–119 (cit. on p. 17).

[Bou18] Boufous, Soufiane, Julie Hatfield, and Raphael Grzebieta: 'The impact of environmental factors on cycling speed on shared paths'. *Accident Analysis & Prevention* (2018), vol. 110: pp. 171–176 (cit. on p. 92).

[Bov85] Bovy, P. and M. Bradley: 'Route Choice Analyzed with Stated-Preference Approaches'. 1985 (cit. on p. 93).

[Bra13] Brajdic, Agata and R. Harle: 'Walk detection and step counting on unconstrained smartphones'. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (2013), vol. (cit. on p. 159).

[Bra08] Brandl, Peter, Clifton Forlines, Daniel Wigdor, Michael Haller, and Chia Shen: 'Combining and Measuring the Benefits of Bimanual Pen and Direct-touch Interaction on Horizontal Interfaces'. *Proceedings of the Working Conference on Advanced Visual Interfaces.* AVI '08. Napoli, Italy: ACM, 2008: pp. 154–161 (cit. on p. 32).

[Bre10] Bresin, Roberto, Anna Dewitt, Stefano Papetti, Marco Civolani, and Federico Fontana: 'Expressive sonification of footstep sounds'. May 2010 (cit. on p. 156).

[Bre20] Bressgott, Timon: 'Strassenprofilschäztung für Fahrräder mit Audio- und Bewegungssignalen unter Verwendung Neuronaler Netze'. Master Thesis. Leibniz Universität Hannover, 2020 (cit. on pp. k, 91).

[Bro12] Broach, Joseph, Jennifer Dill, and John Gliebe: 'Where do cyclists ride? A route choice model developed with revealed preference GPS data'. *Transportation Research Part A: Policy and Practice* (2012), vol. 46(10): pp. 1730–1740 (cit. on p. 93).

[Bro08] Brodie, Matthew, Alan Walmsley, and Wyatt Page: 'Fusion motion capture: a prototype system using inertial measurement units and GPS for the biomechanical analysis of ski racing'. *Sports Technology* (2008), vol. 1(1): pp. 17–28 (cit. on p. 95).

[Bro14] Brooks, Bennett M, Brigitte Schulte-Fortkamp, Kay S Voigt, and Alex U Case: 'Exploring our sonic environment through soundscape research & theory'. *Acoustics today* (2014), vol. 10(1): pp. 30–40 (cit. on p. 174).

[Bru04] Brumm, Henrik: 'The impact of environmental noise on song amplitude in a territorial bird'. *Journal of Animal Ecology* (2004), vol. 73(3): pp. 434–440 (cit. on p. 154).

[Bui13] Buitinck, Lars, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux: 'API design for machine learning software: experiences from the scikit-learn project'. *ECML PKDD Workshop: Languages for Data Mining and Machine Learning.* 2013: pp. 108–122 (cit. on p. 100).

[Bun99] Bunke, H., T. Von Siebenthal, T. Yamasaki, and M. Schenkel: 'Online handwriting data acquisition using a video camera'. *Document Analysis and Recognition, 1999. ICDAR '99. Proceedings of the Fifth International Conference on*. 1999: pp. 573–576 (cit. on p. 33).

[Bur18a] Buriro, Attaullah, Bruno Crispo, Mojtaba Eskandri, Sandeep Gupta, Athar Mahboob, and Rutger Van Acker: 'S nap a uth: a gesture-based unobtrusive smartwatch user authentication scheme'. *International Workshop on Emerging Technologies for Authorization and Authentication*. Springer. 2018: pp. 30–37 (cit. on p. 55).

[Bur18b] Buriro, Attaullah, Rutger Van Acker, Bruno Crispo, and Athar Mahboob: 'Airsign: A gesture-based smartwatch user authentication'. *2018 International Carnahan Conference on Security Technology (ICCST)*. IEEE. 2018: pp. 1–5 (cit. on p. 55).

[But08] Butler, Alex, Shahram Izadi, and Steve Hodges: 'SideSight: Multi-"Touch" Interaction Around Small Devices'. *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*. UIST '08. Monterey, CA, USA: ACM, 2008: pp. 201–204 (cit. on p. 72).

[But19] Butterbaugh, Samantha and Monica Gordon-Pershey: 'The Usefulness of Active Noise Cancelling Headphones when Listening to Music in the Presence of Ambient Noise'. *eHearsay* (2019), vol. 9(2): p. 17 (cit. on pp. 154, 156).

[But05] Butz, Andreas and Ralf Jung: 'Seamless User Notification in Ambient Soundscapes'. *Proceedings of the 10th International Conference on Intelligent User Interfaces*. IUI '05. San Diego, California, USA: Association for Computing Machinery, 2005: pp. 320–322 (cit. on pp. 155–157).

[Bux21] Buxton, Rachel T., Amber L. Pearson, Claudia Allou, Kurt Fristrup, and George Wittemyer: 'A synthesis of health benefits of natural sounds and their distribution in national parks'. *Proceedings of the National Academy of Sciences* (2021), vol. 118(14) (cit. on pp. 155, 156).

[C4020] C40 Cities Climate Leadership Group, Inc.: *Reducing climate change impacts on walking and cycling*. 2020. url: https://www.c40knowledgehub.org/s/article/Reducing-climate-change-impacts-on-walking-and-cycling?language=en_US (visited on 12/27/2021) (cit. on p. 94).

[Cam13] Camacho, Juan Manuel and Victor Sosa: 'Alternative method to calculate the magnetic field of permanent magnets with azimuthal symmetry'. *Revista mexicana de fisica E* (June 2013), vol. 59: pp. 8–17 (cit. on pp. 72, 73).

[Cam18] Camomilla, Valentina, Elena Bergamini, Silvia Fantozzi, and Giuseppe Vannozzi: 'Trends supporting the in-field use of wearable inertial sensors for sport performance evaluation: A systematic review'. *Sensors* (2018), vol. 18(3): p. 873 (cit. on p. 92).

[Cas14] Cassano, John J.: 'Weather Bike: A Bicycle-Based Weather Station for Observing Local Temperature Variations'. *Bulletin of the American Meteorological Society* (2014), vol. 95(2): pp. 205–209 (cit. on pp. 92, 94).

[Cer16] Cerwén, Gunnar, Eja Pedersen, and Anna-Maria Pálsdóttir: 'The role of soundscape in nature-based rehabilitation: A patient perspective'. *International journal of environmental research and public health* (2016), vol. 13(12): p. 1229 (cit. on pp. 155, 156).

[Cha15] Chadwick, Alice C and RW Kentridge: 'The perception of gloss: A review'. *Vision research* (2015), vol. 109: pp. 221–235 (cit. on p. 11).

[Cha02] Cham, Rakié and Mark S Redfern: 'Changes in gait when anticipating slippery floors'. *Gait & posture* (2002), vol. 15(2): pp. 159–171 (cit. on p. 2).

[Cha13] Chan, Louis K.H. and William G. Hayward: 'Visual search'. *Wiley Interdisciplinary Reviews: Cognitive Science* (2013), vol. 4(4): pp. 415–429 (cit. on p. 133).

[Cha17a] Chang, Chia-Ming, Koki Toda, Daisuke Sakamoto, and Takeo Igarashi: 'Eyes on a Car: An Interface Design for Communication between an Autonomous Car and a Pedestrian'. AutomotiveUI '17. Oldenburg, Germany: Association for Computing Machinery, 2017: pp. 65–73 (cit. on pp. 157, 176–178).

[Cha17b] Changhong: *Changhong H2: A Phone with infrared vision.* https://www.cnet.com/reviews/changhong-h2-preview. Accessed: 2021-02-20. 2017 (cit. on p. 117).

[Cha94] Chapman, C Elaine: 'Active versus passive touch: factors influencing the transmission of somatosensory signals to primary somatosensory cortex'. *Canadian journal of physiology and pharmacology* (1994), vol. 72(5): pp. 558–570 (cit. on p. 15).

[Cha09] Chapman, Christine Elaine: 'Active Touch'. *Encyclopedia of Neuroscience*. Ed. by Binder, Marc D., Nobutaka Hirokawa, and Uwe Windhorst. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009: pp. 35–41 (cit. on p. 15).

[Cha16] Chatzidimitris, Thomas, Damianos Gavalas, and Despina Michael: 'SoundPacman: Audio augmented reality in location-based games'. *2016 18th Mediterranean Electrotechnical Conference (MELECON)*. 2016: pp. 1–6 (cit. on p. 155).

[Cha17c] Chatzopoulos, Dimitris, Carlos Bermejo, Zhanpeng Huang, and Pan Hui: 'Mobile augmented reality survey: From where we are to where we go'. *Ieee Access* (2017), vol. 5: pp. 6917–6950 (cit. on p. 24).

[Cha20] Chazette, Larissa, Melanie Busch, Maximilian Schrapel, Kai Korte, and Kurt Schneider: 'Designing Software Transparency: A Multidisciplinary Endeavor'. *Softwaretechnik-Trends* (2020), vol. 40(1): pp. 5–6 (cit. on p. 268).

[Che10] Chen, David M., Sam S. Tsai, Bernd Girod, Cheng-Hsin Hsu, Kyu-Han Kim, and Jatinder Pal Singh: 'Building book inventories using smartphones'. *ACM Multimedia*. 2010 (cit. on p. 134).

[Che11] Chen, Hsiang-Yu, Jaeyoung Park, Steve Dai, and Hong Z Tan: 'Design and evaluation of identifiable key-click signals for mobile devices'. *IEEE Transactions on Haptics* (2011), vol. 4(4): pp. 229–241 (cit. on p. 22).

[Che20] Chen, Mengqi, Jiawei Lin, Yongpan Zou, Rukhsana Ruby, and Kaishun Wu: 'SilentSign: Device-free Handwritten Signature Verification through Acoustic Sensing'. *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2020: pp. 1–10 (cit. on p. 56).

[Che14] Chen, Xiang 'Anthony', Tovi Grossman, and George Fitzmaurice: 'Swipeboard: A Text Entry Technique for Ultra-small Interfaces That Supports Novice to Expert Transitions'. *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*. UIST '14. Honolulu, Hawaii, USA: ACM, 2014: pp. 615–620 (cit. on p. 71).

[Che13] Chen, Ke-Yu, Kent Lyons, Sean White, and Shwetak Patel: 'UTrack: 3D Input Using Two Magnetic Sensors'. UIST '13 (2013), vol.: pp. 237–244 (cit. on p. 72).

[Che16] CHEN, KE-YU, SHWETAK PATEL, and SEAN KELLER: 'Finexus: Tracking Precise Motions of Multiple Fingertips Using Magnetic Sensing'. May 2016: pp. 1504–1514 (cit. on p. 72).

[Che19a] CHEN, YUFENG and BIAO ZHENG: 'What happens after the rare earth crisis: a systematic literature review'. *Sustainability* (2019), vol. 11(5): p. 1288 (cit. on p. 23).

[Che53] CHERRY, E COLIN: 'Some experiments on the recognition of speech, with one and with two ears'. *The Journal of the acoustical society of America* (1953), vol. 25(5): pp. 975–979 (cit. on p. 13).

[Che18] CHEUNG, VICTOR and AUDREY GIROUARD: 'Exploring Around-Device Tangible Interactions for Mobile Devices with a Magnetic Ring'. TEI '18. Stockholm, Sweden: Association for Computing Machinery, 2018: pp. 108–114 (cit. on p. 72).

[Che19b] CHEUNG, VICTOR and AUDREY GIROUARD: 'Tangible Around-Device Interaction Using Rotatory Gestures with a Magnetic Ring'. *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '19. Taipei, Taiwan: Association for Computing Machinery, 2019 (cit. on p. 72).

[Cho06] CHOI, S. d., A. S. LEE, and S. y. LEE: 'On-Line Handwritten Character Recognition with 3D Accelerometer'. *2006 IEEE International Conference on Information Acquisition*. 2006: pp. 845–850 (cit. on p. 33).

[Cho15] CHOLLET, FRANCOIS et al.: *Keras*. 2015. URL: https://github.com/fchollet/keras (cit. on p. 100).

[Cho09] CHOMICZEWSKA, DOROTA, EWA TRZNADEL-BUDŹKO, ANNA KACZOROWSKA, and HELENA ROTSZTEJN: 'The role of Langerhans cells in the skin immune system'. *Polski merkuriusz lekarski: organ Polskiego Towarzystwa Lekarskiego* (2009), vol. 26(153): pp. 173–177 (cit. on p. 14).

[Cho21] CHONG, UIPIL and SHOKHZOD ALIMARDANOV: 'Audio Augmented Reality Using Unity for Marine Tourism'. *Intelligent Human Computer Interaction*. Ed. by SINGH, MADHUSUDAN, DAE-KI KANG, JONG-HA LEE, UMA SHANKER TIWARY, DHANANJAY SINGH, and WAN-YOUNG CHUNG. Cham: Springer International Publishing, 2021: pp. 303–311 (cit. on p. 155).

[Chr99] CHRITIN, V, R BOLOGNESI, and H GUBLER: 'FlowCapt: a new acoustic sensor to measure snowdrift and wind velocity for avalanche forecasting'. *Cold Regions Science and Technology* (1999), vol. 30(1): pp. 125–133 (cit. on p. 95).

[Chu18] Chung, Jungmin, Changhoon Oh, SoHyun Park, and Bongwon Suh: 'PairRing: A Ring-Shaped Rotatable Smartwatch Controller'. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery, 2018: pp. 1–6 (cit. on p. 72).

[Cla94] Clarke, Jeffrey M.: 'Chapter 2 - Neuroanatomy: Brain Structure and Function'. *Neuropsychology*. Ed. by Zaidel, Dahlia W. Handbook of Perception and Cognition. San Diego: Academic Press, 1994: pp. 29–52 (cit. on p. 14).

[Cla16] Clausen, Jens and Liza Steudle: 'Öffentliche Bücherschränke in Hannover Befragungen von PatInnen und NutzerInnen im Auftrag der Landeshauptstadt Hannover'. German. (2016), vol. (cit. on p. 133).

[Cob99] Cobb, Sue VG, Sarah Nichols, Amanda Ramsey, and John R Wilson: 'Virtual reality-induced symptoms and effects (VRISE)'. *Presence: Teleoperators & Virtual Environments* (1999), vol. 8(2): pp. 169–186 (cit. on p. 24).

[Coc21] Cochrane, Karen, Lian Loke, Matthew Leete, Andrew Campbell, and Naseem Ahmadpour: 'Understanding the First Person Experience of Walking Mindfulness Meditation Facilitated by EEG Modulated Interactive Soundscape'. *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. TEI '21. Salzburg, Austria: Association for Computing Machinery, 2021 (cit. on p. 155).

[Coc20] Cochrane, Karen Anne, Lian Loke, Andrew Campbell, Matthew Leete, and Naseem Ahmadpour: 'An Interactive Soundscape to Assist Group Walking Mindfulness Meditation'. *Proceedings of the 7th International Conference on Movement and Computing*. MOCO '20. Jersey City/Virtual, NJ, USA: Association for Computing Machinery, 2020 (cit. on p. 155).

[Coh93] Cohen, M., S. Aoki, and N. Koizumi: 'Augmented audio reality: telepresence/VR hybrid acoustic environments'. *Proceedings of 1993 2nd IEEE International Workshop on Robot and Human Communication*. 1993: pp. 361–364 (cit. on p. 155).

[Cor20] Cornwell, Tara, Jane Woodward, Mengnan Wu, Brennan Jackson, Pamela Souza, Jonathan Siegel, Sumitrajit Dhar, Keith E Gordon, et al.: 'Walking with Ears: Altered Auditory Feedback Impacts Gait Step Length in Older Adults'. *Frontiers in Sports and Active Living* (2020), vol. 2: p. 38 (cit. on pp. 27, 156, 172).

[Cru19] Crum, Poppy: 'Hearables: Here come the: Technology tucked inside your ears will augment your daily life'. *IEEE Spectrum* (2019), vol. 56(5): pp. 38–43 (cit. on p. 156).

[Cur90] Curcio, Christine A, Kenneth R Sloan, Robert E Kalina, and Anita E Hendrickson: 'Human photoreceptor topography'. *Journal of comparative neurology* (1990), vol. 292(4): pp. 497–523 (cit. on p. 10).

[Dal05] Dalal, N. and B. Triggs: 'Histograms of oriented gradients for human detection'. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol. 1. 2005: 886–893 vol. 1 (cit. on pp. 148, 150).

[Dal14] Dalli, Daniele and Matteo Corciolani: 'Gift-giving, sharing and commodity exchange at Bookcrossing.com. New insights from a qualitative analysis'. *Management Decision* (Apr. 2014), vol. 52 (cit. on p. 132).

[Dar15] Darbar, R. and D. Samanta: 'MagiText: Around device magnetic interaction for 3D space text entry in smartphone'. *2015 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. 2015: pp. 1–4 (cit. on p. 72).

[Dar16] Darbar, Rajkumar, Prasanta Kr Sen, and Debasis Samanta: 'PressTact: Side Pressure-Based Input for Smartwatch Interaction'. CHI EA '16. San Jose, California, USA: Association for Computing Machinery, 2016: pp. 2431–2438 (cit. on p. 72).

[Dav08] Davidson, Philip L and Jefferson Y Han: 'Extending 2D object arrangement with pressure-sensitive layering cues'. *Proceedings of the 21st annual ACM symposium on User interface software and technology*. 2008: pp. 87–90 (cit. on p. 20).

[De 21] De Luisa, Luca, Gabriel Emile Hine, Emanuele Maiorana, and Patrizio Campisi: 'In-Air 3D Dynamic Signature Recognition using Haptic Devices'. *2021 IEEE International Workshop on Biometrics and Forensics (IWBF)*. IEEE. 2021: pp. 1–6 (cit. on p. 56).

[dej22] dejure.org: *Beurkundungsgesetz BeurkG §39 Einfache Zeugnisse*. `https://dejure.org/gesetze/BeurkG/39.html`. Accessed: 2022-01-03. 2022 (cit. on pp. 56, 64, 65).

[Dem20] Dementyev, Artem, Alex Olwal, and Richard F. Lyon: 'Haptics with Input: Back-EMF in Linear Resonant Actuators to Enable Touch, Pressure and Environmental Awareness'. *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. UIST

'20. Virtual Event, USA: Association for Computing Machinery, 2020: pp. 420–429 (cit. on p. 118).

[Den01] Denning, Peter J: *The Invisible future: the seamless integration of technology into everyday life*. McGraw-Hill, Inc., 2001 (cit. on p. 16).

[Dia20] Dian, F John, Reza Vahidnia, and Alireza Rahmati: 'Wearables and the Internet of Things (IoT), applications, opportunities, and challenges: A Survey'. *IEEE Access* (2020), vol. 8: pp. 69200–69211 (cit. on p. 92).

[Die01] Dietz, Paul and Darren Leigh: 'DiamondTouch: A Multi-User Touch Technology'. Jan. 2001 (cit. on p. 19).

[Dil17] Dillahunt, Tawanna R., Xinyi Wang, Earnest Wheeler, Hao Fei Cheng, Brent Hecht, and Haiyi Zhu: 'The Sharing Economy in Computing: A Systematic Literature Review'. *Proc. ACM Hum.-Comput. Interact.* (Dec. 2017), vol. 1(CSCW): 38:1–38:26 (cit. on p. 131).

[Din19] Ding, Feng, Dong Wang, Qian Zhang, and Run Zhao: 'ASSV: Handwritten Signature Verification Using Acoustic Signals'. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (Sept. 2019), vol. 3(3) (cit. on p. 56).

[Din10] Dingler, Tilman and Stephen Brewster: 'AudioFeeds: A Mobile Auditory Application for Monitoring Online Activities'. MM '10. Firenze, Italy: Association for Computing Machinery, 2010: pp. 1067–1070 (cit. on pp. 155, 156).

[Dix04] Dix, Alan, Janet Finlay, Gregory D Abowd, and Russell Beale: *Human-computer interaction*. Pearson Education, 2004 (cit. on p. 9).

[Doc21] DocuSign: *Are Electronic Signatures Legal?* https://www.docusign.com/blog/are-electronic-signatures-legal. Accessed: 2022-01-03. 2021 (cit. on p. 54).

[Doe11] Doerschner, Katja, Roland W Fleming, Ozgur Yilmaz, Paul R Schrater, Bruce Hartung, and Daniel Kersten: 'Visual motion and the perception of surface material'. *Current Biology* (2011), vol. 21(23): pp. 2010–2016 (cit. on p. 11).

[Dog20] Dogan, Mustafa Doga, Faraz Faruqi, Andrew Day Churchill, Kenneth Friedman, Leon Cheng, Sriram Subramanian, and Stefanie Mueller: 'G-ID: Identifying 3D Prints Using Slicing Parameters'. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, 2020: pp. 1–13 (cit. on p. 118).

[Dou11] Dourish, Paul and Genevieve Bell: *Divining a digital future: Mess and mythology in ubiquitous computing*. Mit Press, 2011 (cit. on p. 16).

[Dow05] Downs, Rick: 'Using resistive touch screens for human/machine interface'. *Analog Applications Journal Q* (2005), vol. 3: pp. 5–10 (cit. on p. 19).

[Dre20] Drey, Tobias, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio: 'Vrsketchin: Exploring the design space of pen and tablet interaction for 3d sketching in virtual reality'. *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020: pp. 1–14 (cit. on p. 186).

[Dro02] Droste, Stefan, Thomas Jansen, and Ingo Wegener: 'On the analysis of the (1+ 1) evolutionary algorithm'. *Theoretical Computer Science* (2002), vol. 276(1-2): pp. 51–81 (cit. on p. 161).

[Du19] Du, Rong, Paolo Santi, Ming Xiao, Athanasios V. Vasilakos, and Carlo Fischione: 'The Sensable City: A Survey on the Deployment and Management for Smart City Monitoring'. *IEEE Communications Surveys Tutorials* (2019), vol. 21(2): pp. 1533–1560 (cit. on p. 92).

[Dud19] Dudley, John, Hrvoje Benko, Daniel Wigdor, and Per Ola Kristensson: 'Performance envelopes of virtual keyboard text input strategies in virtual reality'. *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2019: pp. 289–300 (cit. on pp. 26, 182).

[Dun07] Dunlop, Mark, Brian Elsey, and Michelle Masters: 'Dynamic visualisation of ski data: A context aware mobile piste map'. Jan. 2007: pp. 375–378 (cit. on p. 94).

[Dun14] Dunlop, Mark D., Andreas Komninos, and Naveen Durga: 'Towards High Quality Text Entry on Smartwatches'. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '14. Toronto, Ontario, Canada: ACM, 2014: pp. 2365–2370 (cit. on p. 71).

[Eck01] Eckel, G.: 'Immersive audio-augmented environments: the LISTEN project'. *Proceedings Fifth International Conference on Information Visualisation*. 2001: pp. 571–573 (cit. on p. 155).

[Ehr12] Ehrgott, Matthias, Judith Y.T. Wang, Andrea Raith, and Chris van Houtte: 'A bi-objective cyclist route choice model'. *Transportation Research Part A: Policy and Practice* (2012), vol. 46(4): pp. 652–663 (cit. on p. 93).

[El 11] El Saddik, Abdulmotaleb, Mauricio Orozco, Mohamad Eid, and Jongeun Cha: *Haptics technologies: Bringing touch to multimedia.* Springer Science & Business Media, 2011 (cit. on p. 9).

[Elm12] Elmqvist, Niklas and Karthik Ramani: 'Extended Multitouch: Recovering Touch Posture and Differentiating Users Using a Depth Camera'. (2012), vol. (cit. on p. 18).

[Eri17] Erica Johnson, CBC News: *Document forgery in financial industry more common than you'd think, past employees say.* https://www.cbc.ca/news/business/financial-industry-employees-forge-documents-more-often-than-you-d-think-1.4138212. Accessed: 2022-01-03. 2017 (cit. on p. 54).

[Eri19] Erickson, Z., N. Luskey, S. Chernova, and C. C. Kemp: 'Classification of Household Materials via Spectroscopy'. *IEEE Robotics and Automation Letters* (2019), vol. 4(2): pp. 700–707 (cit. on p. 118).

[Eri08] Eriksson, Jakob, Lewis Girod, Bret Hull, Ryan Newton, Samuel Madden, and Hari Balakrishnan: 'The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring'. *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*. MobiSys '08. Breckenridge, CO, USA: Association for Computing Machinery, 2008: pp. 29–39 (cit. on pp. 92–94).

[Esh17] Esha Nerurkar Simon Lynen, Sheng Zhao: *System and method for concurrent odometry and mapping.* US Patent 20170336511A1. Nov. 2017 (cit. on pp. 25, 140).

[Etg18] Etgeton, Philipp: 'Erkennung spektroskopischer Charateristika von Oberflächen mit Smartphonekameras'. Master Thesis. Leibniz Universität Hannover, 2018 (cit. on p. 115).

[Eth21a] Ethereum.org: *Decentralized finance (DeFi) - What is DeFI?* https://ethereum.org/en/defi/. Accessed: 2022-01-03. 2021 (cit. on p. 54).

[Eth21b] Ethereum.org, Paul Wackerow: *Introduction to smart contracts.* https://ethereum.org/en/developers/docs/smart-contracts/. Accessed: 2022-01-03. 2021 (cit. on p. 54).

[Fan16] Fan, Liming, Chong Kang, Xiaojun Zhang, Quan Zheng, and Ming Wang: 'An efficient method for tracking a magnetic target using scalar magnetometer array'. *SpringerPlus* (Dec. 2016), vol. 5 (cit. on pp. 72, 73).

[Fan05] FANG, PING, ZHONGCHENG WU, FEI SHEN, YUNJIAN GE, and BING FANG: 'Improved DTW Algorithm for Online Signature Verification Based on Writing Forces'. *Proceedings of the 2005 International Conference on Advances in Intelligent Computing - Volume Part I*. ICIC'05. Hefei, China: Springer-Verlag, 2005: pp. 631–640 (cit. on p. 55).

[Fed18] FEDOSOV, ANTON, JEREMIAS ALBANO, and MARC LANGHEINRICH: 'Supporting the Design of Sharing Economy Services: Learning from Technology-mediated Sharing Practices of Both Digital and Physical Artifacts'. *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*. NordiCHI '18. Oslo, Norway: ACM, 2018: pp. 323–337 (cit. on p. 132).

[Fed15] FEDOSOV, ANTON and MARC LANGHEINRICH: 'From Start to Finish: Understanding Group Sharing Behavior in a Backcountry Skiing Community'. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '15. Copenhagen, Denmark: Association for Computing Machinery, 2015: pp. 758–765 (cit. on p. 94).

[Fed16] FEDOSOV, ANTON, EVANGELOS NIFORATOS, IVAN ELHART, TESEO SCHNEIDER, DMITRY ANISIMOV, and MARC LANGHEINRICH: 'Design and Evaluation of a Wearable AR System for Sharing Personalized Content on Ski Resort Maps'. Dec. 2016 (cit. on p. 94).

[Feh17] FEHER, JOSEPH J: *Quantitative human physiology: an introduction*. Academic press, 2017 (cit. on pp. 9–13, 16).

[Fei93] FEINER, STEVEN, BLAIR MACINTYRE, and DORÉE SELIGMANN: 'Knowledge-Based Augmented Reality'. *Commun. ACM* (July 1993), vol. 36(7): pp. 53–62 (cit. on pp. 23, 28).

[Fei20] FEIZI, AHMAD, JUN-SEOK OH, VALERIAN KWIGIZILE, and SHINHYE JOO: 'Cycling environment analysis by bicyclists' skill levels using instrumented probe bicycle (IPB)'. *International Journal of Sustainable Transportation* (2020), vol. 14(9): pp. 722–732 (cit. on p. 93).

[Fie07] FIERREZ, JULIAN, JAVIER ORTEGA-GARCIA, DANIEL RAMOS, and JOAQUIN GONZALEZ-RODRIGUEZ: 'HMM-based on-line signature verification: Feature extraction and signature modeling'. *Pattern Recognition Letters* (Dec. 2007), vol. 28: pp. 2325–2334 (cit. on p. 55).

[Fie21] FIETS, SNUFFEL: *Towards a cycling measurement network*. 2021. URL: https://snuffelfiets.nl/ (visited on 12/27/2021) (cit. on pp. 92, 94).

[Fis81] Fischler, Martin A and Robert C Bolles: 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography'. *Communications of the ACM* (1981), vol. 24(6): pp. 381–395 (cit. on p. 26).

[Fit93] Fitzmaurice, George W: 'Situated information spaces and spatially aware palmtop computers'. *Communications of the ACM* (1993), vol. 36(7): pp. 39–49 (cit. on p. 23).

[Fit99] Fitzpatrick, Richard C., Daniel L. Wardman, and Janet L. Taylor: 'Effects of galvanic vestibular stimulation during human walking'. *The Journal of Physiology* (1999), vol. 517 (cit. on p. 13).

[Fou13] Foulke, Emerson: 'Braille'. *The psychology of touch*. Psychology Press, 2013: pp. 231–246 (cit. on p. 2).

[Fur21] Furst, R Terry and Douglas N Evans: 'Emotional reactions to loss and recovery of a smartphone: implications for habituation'. *Journal of Technology in Behavioral Science* (2021), vol. 6(3): pp. 527–534 (cit. on p. 126).

[Gam80] Gamble, DJ: 'The handwriting of identical twins'. *Canadian Society of Forensic Science Journal* (1980), vol. 13(1): pp. 11–30 (cit. on pp. 54, 66).

[Gam14] Gamper, H.: 'Enabling technologies for audio augmented reality systems'. 2014 (cit. on p. 155).

[ges22a] gesetze-im-internet.de: *Bürgerliches Gesetzbuch (BGB) §126 Schriftform*. http://www.gesetze-im-internet.de/bgb/__126.html. Accessed: 2022-06-14. 2022 (cit. on p. 54).

[ges22b] gesetze-im-internet.de: *Bürgerliches Gesetzbuch (BGB) §2247 Eigenhändiges Testament*. https://www.gesetze-im-internet.de/bgb/__2247.html. Accessed: 2022-06-14. 2022 (cit. on p. 54).

[Ges20] Gesslein, Travis, Verena Biener, Philipp Gagel, Daniel Schneider, Per Ola Kristensson, Eyal Ofek, Michel Pahud, and Jens Grubert: 'Pen-based interaction with spreadsheets in mobile virtual reality'. *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2020: pp. 361–373 (cit. on p. 186).

[Giæ98] Giæver, T, L Øvstedal, and T Lindland: 'Utforming av sykkelanlegg–Intervju og atferdsundersøkelser'. *SINTEF rapport STF22 A* (1998), vol. 97615 (cit. on p. 94).

[Gia17] Giancaspro, Mark: 'Is a smart contract really a smart idea? Insights from a legal perspective'. *Computer Law & Security Review* (2017), vol. 33(6): pp. 825–835 (cit. on p. 54).

[Gib94] Gibson, Bradley S and Mary A Peterson: 'Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cuing paradigm.' *Journal of Experimental Psychology: Human Perception and Performance* (1994), vol. 20(2): p. 299 (cit. on p. 8).

[Gib62] Gibson, James J: 'Observations on active touch.' *Psychological review* (1962), vol. 69(6): p. 477 (cit. on p. 15).

[Gil17] Gil, Hyunjae, DoYoung Lee, Seunggyu Im, and Ian Oakley: 'TriTap: Identifying Finger Touches on Smartwatches'. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017: pp. 3879–3890 (cit. on p. 71).

[Gio12] Gioi, Rafael Grompone von, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall: 'LSD: a Line Segment Detector'. *IPOL Journal* (2012), vol. 2: pp. 35–55 (cit. on pp. 134, 136).

[Gir11] Girod, B., V. Chandrasekhar, R. Grzeszczuk, and Y. A. Reznik: 'Mobile Visual Search: Architectures, Technologies, and the Emerging MPEG Standard'. *IEEE MultiMedia* (2011), vol. 18(3): pp. 86–94 (cit. on p. 133).

[Goi18] Goicoechea-Telleria, Ines, Raul Sanchez-Reillo, Judith Liu-Jimenez, and Ramon Blanco-Gonzalo: 'Attack potential evaluation in desktop and smartphone fingerprint sensors: can they be attacked by anyone?' *Wireless Communications and Mobile Computing* (2018), vol. 2018 (cit. on p. 66).

[Gol09] Goldstein, E Bruce: *Encyclopedia of perception*. Sage, 2009 (cit. on pp. 9, 14, 16).

[Gol17] Goldstein, E. Bruce and James R. Brockmole: *Sensation and Perception*. Cengage Learning, 2017 (cit. on pp. 8, 9, 11, 13).

[Gol99] Goldstein, Mikael and Didier Chincholle: 'Finger-joint gesture wearable keypad'. *second workshop on human computer interaction with mobile devices*. 1999 (cit. on p. 25).

[Gol95] Golledge, R.: 'Defining the Criteria Used in Path Selection'. 1995 (cit. on p. 93).

[Gom20] Gomez-Andres, Alba, Jennifer Grau-Sánchez, Esther Duarte, Antoni Rodriguez-Fornells, and Ana Tajadura-Jiménez: 'Enriching footsteps sounds in gait rehabilitation in chronic stroke patients: a pilot study'. *Annals of the New York Academy of Sciences* (2020), vol. 1467(1): pp. 48–59 (cit. on pp. 13, 27, 156).

[Gon20] Gong, Jun, Aakar Gupta, and Hrvoje Benko: 'Acustico: Surface tap detection and localization using wrist-based acoustic tdoa sensing'. *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 2020: pp. 406–419 (cit. on p. 25).

[Gon18] Gong, Jun, Zheer Xu, Qifan Guo, Teddy Seyed, Xiang 'Anthony' Chen, Xiaojun Bi, and Xing-Dong Yang: 'WrisText: One-handed Text Entry on Smartwatch Using Wrist Gestures'. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: ACM, 2018: 181:1–181:14 (cit. on p. 72).

[Gon16] Gong, Jun, Xing-Dong Yang, and Pourang Irani: 'WristWhirl: One-handed Continuous Smartwatch Input using Wrist Gestures'. Oct. 2016: pp. 861–872 (cit. on p. 72).

[Goo18] Google: *Google Android API Guide*. Accessed: 2021-02-02. 2018 (cit. on pp. 118, 127, 128, 135).

[Goo19] Google: *Google Lens*. https://lens.google.com. 2019 (cit. on p. 134).

[Goo20] Google: *Material Icons*. Website. https://material.io/resources/icons/?; Retrieved 1st. May 2020. 2020 (cit. on p. 85).

[Gor16] Gordon, Mitchell, Tom Ouyang, and Shumin Zhai: 'WatchWriter: Tap and Gesture Typing on a Smartwatch Miniature Keyboard with Statistical Decoding'. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, 2016: pp. 3817–3821 (cit. on p. 71).

[Gra18] Grannemann, Dennis: 'Gesture Recognition with Audio and Motion Data from Digital Pens'. Bachelor Thesis. Leibniz Universität Hannover, 2018 (cit. on pp. k, 53).

[Gri13] Griechisch, Erika, M Imran Malik, and Marcus Liwicki: 'Online signature verification using accelerometer and gyroscope'. *International Graphonomics Society* (2013), vol. (cit. on p. 56).

[Gri17] Griswold-Steiner, Isaac, Richard Matovu, and Abdul Serwadda: 'Handwriting watcher: A mechanism for smartwatch-driven handwriting authentication'. *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE. 2017: pp. 216–224 (cit. on p. 56).

[Gro16] Groh, B. H., M. Fleckenstein, and B. M. Eskofier: 'Wearable trick classification in freestyle snowboarding'. *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 2016: pp. 89–93 (cit. on p. 95).

[Grø07] Grønbæk, Kaj, Ole S Iversen, Karen Johanne Kortbek, Kaspar Rosengreen Nielsen, and Louise Aagaard: 'IGameFloor: a platform for co-located collaborative games'. *Proceedings of the international conference on Advances in computer entertainment technology*. 2007: pp. 64–71 (cit. on p. 20).

[Gru09] Gruber, Christian, Thiemo Gruber, Sebastian Krinninger, and Bernhard Sick: 'Online signature verification with support vector machines based on LCSS kernel functions'. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* (2009), vol. 40(4): pp. 1088–1100 (cit. on p. 55).

[Gru01] Grudin, Jonathan: 'Integrating paper and digital information on EnhancedDesk: a method for realtime finger tracking on an augmented desk system'. *ACM Transactions on Computer-Human Interaction (TOCHI)* (2001), vol. 8(4): pp. 307–322 (cit. on p. 26).

[Gru18] Grundhöfer, Anselm and Daisuke Iwai: 'Recent advances in projection mapping algorithms, hardware and applications'. *Computer Graphics Forum*. Vol. 37. 2. Wiley Online Library. 2018: pp. 653–675 (cit. on p. 24).

[Gru08] Grunwald, Martin and Matthias John: 'German pioneers of research into human haptic perception'. *Human haptic perception: basics and applications*. Springer, 2008: pp. 15–39 (cit. on p. 9).

[Gue18] Gueorguiev, David, Anis Kaci, Michel Amberg, Frédéric Giraud, and Betty Lemaire-Semail: 'Travelling ultrasonic wave enhances keyclick sensation'. *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer. 2018: pp. 302–312 (cit. on p. 22).

[Gue15] Guerbai, Yasmine, Youcef Chibani, and Bilal Hadjadji: 'The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters'. *Pattern Recognition* (2015), vol. 48(1): pp. 103–113 (cit. on p. 55).

[Gue21] Guerra-Segura, Elyoenai, Aysse Ortega-Pérez, and Carlos M Travieso: 'In-air signature verification system using Leap Motion'. *Expert Systems with Applications* (2021), vol. 165: p. 113797 (cit. on p. 55).

[Gup16] Gupta, Aakar and Ravin Balakrishnan: 'DualKey: Miniature Screen Text Entry via Finger Identification'. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: ACM, 2016: pp. 59–70 (cit. on p. 72).

[Gup18] Gupta, Aakar, Jiushan Yang, and Ravin Balakrishnan: 'Asterisk and Obelisk: Motion Codes for Passive Tagging'. *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. UIST '18. Berlin, Germany: Association for Computing Machinery, 2018: pp. 725–736 (cit. on p. 72).

[GuS21] GuS Kommunikation and Maximilian Schrapel: 'MINT-Macher im Gespräch: Maximilian Schrapel'. *TTS: Treffpunkt Technik in der Schule* (2021), vol. 2 (cit. on p. 269).

[Haa18] Haas, Gabriel, Evgeny Stemasov, and Enrico Rukzio: 'Can't You Hear Me? Investigating Personal Soundscape Curation'. *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*. MUM 2018. Cairo, Egypt: Association for Computing Machinery, 2018: pp. 59–69 (cit. on p. 154).

[Haf17] Hafemann, Luiz G, Robert Sabourin, and Luiz S Oliveira: 'Offline handwritten signature verification - literature review'. *2017 seventh international conference on image processing theory, tools and applications (IPTA)*. IEEE. 2017: pp. 1–8 (cit. on p. 55).

[Haj16] Hajinejad, Nassrin, Barbara Grüter, Licinio Roque, and Simon Bogutzky: 'GangKlang: Facilitating a movement-oriented walking experience through sonic interaction'. *Proceedings of the Audio Mostly 2016*. 2016: pp. 202–208 (cit. on p. 156).

[Hal17] Haladjian, Juan, Maximilian Reif, and Bernd Bruegge: 'VIHapp: a wearable system to support blind skiing'. Sept. 2017: pp. 1033–1037 (cit. on p. 95).

[Han20] Han, H., H. Jang, and S. W. Yoon: 'Novel Wearable Monitoring System of Forward Head Posture Assisted by Magnet-Magnetometer Pair and Machine Learning'. *IEEE Sensors Journal* (2020), vol. 20(7): pp. 3838–3848 (cit. on p. 72).

[Han05] Han, Jefferson Y: 'Low-cost multi-touch sensing through frustrated total internal reflection'. *Proceedings of the 18th annual ACM symposium on User interface software and technology*. 2005: pp. 115–118 (cit. on p. 20).

[Han11] Hanh, Thich Nhat: *The long road turns to joy: A guide to walking meditation*. Parallax Press, 2011 (cit. on p. 175).

[Han85] Hanh, Thick Nhat: 'A guide to walking meditation'. *Fellowship* (1985), vol. 51(10-11): p. 23 (cit. on pp. 13, 154, 158).

[Han16] Hannover, Stadt: *10 Jahre Offene Buecherschraenke in Hannover - Eine Erfolgsgeschichte*. `https://e-government.hannover-stadt.de`. 2016 (cit. on pp. 133, 149).

[Hap21] Happe, Janko: 'Augmenting Footstep Sounds with Noise Cancelling Headphones in Public Places'. Master Thesis. Leibniz Universität Hannover, 2021 (cit. on pp. k, 153).

[Har78] Harris, Frederic J.: 'On the Use of Windows for Harmonic Analysis With the Discrete Fourier Transform'. (Feb. 1978), vol. 66: pp. 51–83 (cit. on p. 38).

[Har11] Harrison, Chris, Hrvoje Benko, and Andrew D Wilson: 'OmniTouch: wearable multitouch interaction everywhere'. *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 2011: pp. 441–450 (cit. on pp. 25, 26).

[Har09] Harrison, Chris and Scott Hudson: 'Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices.' Jan. 2009: pp. 121–124 (cit. on p. 72).

[Har08a] Harrison, Chris and Scott E. Hudson: 'Lightweight Material Detection for Placement-aware Mobile Computing'. *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*. UIST '08. Monterey, CA, USA: ACM, 2008: pp. 279–282 (cit. on pp. 117, 118).

[Har08b] Harrison, Chris and Scott E. Hudson: 'Scratch Input: Creating Large, Inexpensive, Unpowered and Mobile Finger Input Surfaces'. *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*. UIST '08. Monterey, CA, USA: ACM, 2008: pp. 205–208 (cit. on pp. 4, 19, 34, 58).

[Har10] Harrison, Chris, Desney Tan, and Dan Morris: 'Skinput: appropriating the body as an input surface'. *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010: pp. 453–462 (cit. on pp. 24, 25).

[Has12] Hasegawa, Shoichi, Seiichiro Ishijima, Fumihiro Kato, Hironori Mitake, and Makoto Sato: 'Realtime Sonification of the Center of Gravity for Skiing'. *Proceedings of the 3rd Augmented Human International Conference*. AH '12. Megève, France: Association for Computing Machinery, 2012 (cit. on p. 95).

[Hee11] Heesch, Kristiann C, Jan Garrard, and Shannon Sahlqvist: 'Incidence, severity and correlates of bicycling injuries in a sample of cyclists in Queensland, Australia'. *Accident Analysis & Prevention* (2011), vol. 43(6): pp. 2085–2092 (cit. on p. 92).

[Hel20] Heller, Florian, Jelco Adamczyk, and Kris Luyten: 'Attracktion: Field Evaluation of Multi-Track Audio as Unobtrusive Cues for Pedestrian Navigation'. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '20. Oldenburg, Germany: Association for Computing Machinery, 2020 (cit. on pp. 155, 175).

[Hel15] Heller, Florian and Jan Borchers: 'Audioscope: Smartphones as directional microphones in mobile audio augmented reality systems'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015: pp. 949–952 (cit. on p. 155).

[Hel14] Heller, Florian and Jan Borchers: 'AudioTorch: using a smartphone as directional microphone in virtual audio spaces'. *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. 2014: pp. 483–488 (cit. on p. 155).

[Hel18] Heller, Florian and Johannes Schöning: 'Navigatone: Seamlessly embedding navigation cues in mobile music listening'. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018: pp. 1–7 (cit. on p. 155).

[Hen04] Henderson, T. C., E. Grant, K. Luthy, and J. Cintron: 'Snow monitoring with sensor networks'. *29th Annual IEEE International Conference on Local Computer Networks*. 2004: pp. 558–559 (cit. on p. 95).

[Her12] Hermann, Thomas, Anselm Venezian Nehls, Florian Eitel, Tarik Barri, and Marcus Gammel: 'Tweetscapes - Real-Time Sonification of Twitter Data Streams for Radio Broadcasting'. 2012 (cit. on pp. 155, 156).

[Her19] Herzog, Florian: 'Implementation und Evaluation eines magnetischen Zeige- gerätes für Smartwatches'. Bachelor Thesis. Leibniz Universität Hannover, 2019 (cit. on pp. k, 69).

[Hil92] Hilton, Ordway: *Scientific examination of questioned documents*. CRC press, 1992 (cit. on pp. 54, 65).

[Hof10] Hofer, Ramon Linus and Andreas Kunz: 'Digisketch: taming anoto technology on LCDs'. *Proceedings of the 2nd ACM SIGCHI symposium on Engineering interactive computing systems*. 2010: pp. 103–108 (cit. on p. 21).

[Hol10a] Holleczek, T., A. Rüegg, H. Harms, and G. Tröster: 'Textile pressure sensors for sports applications'. *SENSORS, 2010 IEEE*. 2010: pp. 732–737 (cit. on p. 95).

[Hol10b] Holleczek, Thomas, Jona Schoch, Bert Arnrich, and G. Tröster: 'Recognizing turns and other snowboarding activities with a gyroscope'. Nov. 2010: pp. 1–8 (cit. on p. 95).

[Hol09] Holleczek, Thomas, Christoph Zysset, Bert Arnrich, Daniel Roggen, and Gerhard Tröster: 'Towards an Interactive Snowboarding Assistance System'. Sept. 2009: pp. 147–148 (cit. on pp. 95, 97).

[Höl12] Hölzel, Christin, Franz Höchtl, and Veit Senner: 'Cycling comfort on different road surfaces'. *Procedia Engineering* (2012), vol. 34. ENGINEERING OF SPORT CONFERENCE 2012: pp. 479–484 (cit. on pp. 92, 94).

[Hon15] Hong, Jonggi, Seongkook Heo, Poika Isokoski, and Geehyuk Lee: 'SplitBoard: A Simple Split Soft Keyboard for Wristwatch-sized Touch Screens'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, 2015: pp. 1233–1236 (cit. on p. 71).

[Hon17] Hong, Joo Young, Jianjun He, Bhan Lam, Rishabh Gupta, and Woon-Seng Gan: 'Spatial Audio for Soundscape Design: Recording and Reproduction'. *Applied Sciences* (2017), vol. 7(6) (cit. on p. 155).

[Hoo04] Hook, Christian, Juergen Kempf, and Georg Scharfenberg: 'A novel digitizing pen for the analysis of pen pressure and inclination in handwriting biometrics'. *International Workshop on Biometric Authentication*. Springer. 2004: pp. 283–294 (cit. on p. 67).

[Hoo03] Hook, Christian, Juergen Kempf, and Georg Scharfenberg: 'New Pen Device for Biometrical 3D Pressure Analysis of Handwritten Characters, Words and Signatures'. *Proceedings of the 2003 ACM SIGMM Workshop on Biometrics Methods and Applications*. WBMA '03. Berkley, California: Association for Computing Machinery, 2003: pp. 38–44 (cit. on p. 56).

[Hop19] Hoppe, Matthias, Jakob Karolus, Felix Dietz, Paweł W. Woźniak, Albrecht Schmidt, and Tonja-Katrin Machulla: 'VRsneaky: Increasing Presence in VR Through Gait-Aware Auditory Feedback'. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2019: pp. 1–9 (cit. on pp. 13, 27, 156).

[How17] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam: 'MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications'. *CoRR* (2017), vol. abs/1704.04861 (cit. on pp. 135, 136).

[Hsu15] Hsu, Yu-Liang, Cheng-Ling Chu, Yi-Ju Tsai, and Jeen-Shing Wang: 'An Inertial Pen With Dynamic Time Warping Recognizer for Handwriting and Gesture Recognition'. *IEEE Sensors Journal* (2015), vol. 15(1): pp. 154–163 (cit. on pp. 4, 56, 58).

[Hub99] Huber, Roy A and Alfred M Headrick: *Handwriting identification: facts and fundamentals*. CRC press, 1999 (cit. on pp. 54, 65).

[Hue16] Huebner, Gesche, D.T. Shipworth, Stephanie Gauthier, Christoph Witzel, P Raynham, and W Chan: 'Saving energy with light? Experimental studies assessing the impact of colour temperature on thermal comfort'. *Energy Research & Social Science* (May 2016), vol. 15: pp. 45–57 (cit. on p. 119).

[Hwa13a] Hwang, Sungjae, Myungwook Ahn, and Kwang-yun Wohn: 'MagGetz: Customizable Passive Tangible Controllers on and around Conventional Mobile Devices'. *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*. UIST '13. St. Andrews, Scotland, United Kingdom: Association for Computing Machinery, 2013: pp. 411–416 (cit. on p. 72).

[Hwa13b] Hwang, Sungjae, Andrea Bianchi, Myungwook Ahn, and Kwangyun Wohn: 'MagPen: Magnetically Driven Pen Interactions on and around Conventional Smartphones'. *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '13. Munich, Germany: Association for Computing Machinery, 2013: pp. 412–415 (cit. on pp. 21, 72).

[Hwa12] Hwang, Sungjae, Andrea Bianchi, and Kwangyun Wohn: 'MicPen: Pressure-Sensitive Pen Interaction Using Microphone with Standard Touchscreen'. *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '12. Austin, Texas, USA: Association for Computing Machinery, 2012: pp. 1847–1852 (cit. on pp. 21, 34, 40, 64).

[Imp08] Impedovo, Donato and Giuseppe Pirlo: 'Automatic Signature Verification: The State of the Art'. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* (2008), vol. 38(5): pp. 609–635 (cit. on p. 55).

[Inc19] Inc., Google: *ML-Kit: Machine learning for mobile developers.* https://developers.google.com/ml-kit/. 2019 (cit. on pp. 135, 139).

[Inc18] Inc., Google: *Sceneform SDK for Android.* https://github.com/google-ar/sceneform-android-sdk. 2018 (cit. on pp. 135, 140).

[Inc] Incorporated, Adobe Systems: *Mixamo: Animate 3D characters for games, film, and more.* https://www.mixamo.com/. Accessed: 2022-02-24 (cit. on p. 176).

[Ira13] Iranmanesh, Vahab, Sharifah Mumtazah Syed Ahmad, Wan Azizun Wan Adnan, Fahad Layth Malallah, and Salman Yussof: 'Online signature verification using neural network and pearson correlation features'. *2013 IEEE conference on open systems (ICOS).* IEEE. 2013: pp. 18–21 (cit. on p. 55).

[Isi03] Ising, Hartmut and B Kruppa: 'Health effects caused by noise: Evidence in the literature from the past 25 years'. *Noise & health* (Nov. 2003), vol. 6: pp. 5–13 (cit. on pp. 154, 155).

[Isr12] Israr, Ali, Olivier Bau, Seung-Chan Kim, and Ivan Poupyrev: 'Tactile feedback on flat surfaces for the visually impaired'. *CHI'12 Extended Abstracts on Human Factors in Computing Systems.* 2012: pp. 1571–1576 (cit. on p. 23).

[Ito21] Itoh, Yuta, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski: 'Towards indistinguishable augmented reality: A survey on optical see-through head-mounted displays'. *ACM Computing Surveys (CSUR)* (2021), vol. 54(6): pp. 1–36 (cit. on pp. 24, 185).

[Its15] Itseez: *Open Source Computer Vision Library.* https://github.com/itseez/opencv. 2015 (cit. on p. 135).

[Iwa16] Iwana, Brian Kenji, Syed Tahseen Raza Rizvi, Sheraz Ahmed, Andreas Dengel, and Seiichi Uchida: 'Judging a Book by its Cover'. *arXiv preprint arXiv:1610.09204* (2016), vol. (cit. on p. 141).

[Iwa10] Iwata, Kazumasa, Koichi Kise, Masakazu Iwamura, Seiichi Uchida, and Shinichiro Omachi: 'Tracking and retrieval of pen tip positions for an intelligent camera pen'. *2010 12th International Conference on Frontiers in Handwriting Recognition.* IEEE. 2010: pp. 277–282 (cit. on p. 21).

[Jan17] Janus, Joanna M, Ryan FL O'Shaughnessy, Catherine A Harwood, and Tania Maffucci: 'Phosphoinositide 3-kinase-dependent signalling pathways in cutaneous squamous cell carcinomas'. *Cancers* (2017), vol. 9(7): p. 86 (cit. on p. 14).

[Jay13] Jayalath, Sampath and Nimsiri Abhayasinghe: 'A gyroscopic data based pedometer algorithm'. *2013 8th International Conference on Computer Science Education*. 2013: pp. 551–555 (cit. on p. 159).

[Jer09] Jerald, Jason and Mary Whitton: 'Relating scene-motion thresholds to latency thresholds for head-mounted displays'. *2009 IEEE virtual reality conference*. IEEE. 2009: pp. 211–218 (cit. on p. 24).

[Joh65] Johnson, Eric Arthur: 'Touch display–a novel input/output device for computers'. *Electronics Letters* (1965), vol. 1(8): pp. 219–220 (cit. on p. 17).

[Joh17] Johnson, I., J. Henderson, C. Perry, J. Schöning, and B. Hecht: 'Beautiful... but at What Cost? An Examination of Externalities in Geographic Vehicle Routing'. (June 2017), vol. 1(2) (cit. on p. 93).

[Jus01] Justino, Edson JR, Flávio Bortolozzi, and Robert Sabourin: 'Off-line signature verification using HMM for random, simple and skilled forgeries'. *Proceedings of Sixth International Conference on Document Analysis and Recognition*. IEEE. 2001: pp. 1031–1034 (cit. on p. 55).

[Kab82] Kabat-Zinn, Jon: 'An outpatient program in behavioral medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results'. *General Hospital Psychiatry* (1982), vol. 4(1): pp. 33–47 (cit. on pp. 13, 158).

[Kab13] Kabat-Zinn, Jon: *Full catastrophe living, revised edition: how to cope with stress, pain and illness using mindfulness meditation*. Hachette uK, 2013 (cit. on pp. 13, 158).

[Kab17] Kabat-Zinn, Jon: 'Walking Meditations'. *Mindfulness* (Feb. 2017), vol. 8(1): pp. 249–250 (cit. on p. 154).

[Kaj14] Kajimoto, Hiroyuki, Masaki Suzuki, and Yonezo Kanno: 'HamsaTouch: Tactile Vision Substitution with Smartphone and Electro-Tactile Display'. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014: pp. 1273–1278 (cit. on p. 27).

[Kan16] Kang, Jian, Francesco Aletta, Truls Gjestland, Lex Brown, Dick Botteldooren, Brigitte Schulte-Fortkamp, Peter Lercher, Irene van Kamp, Klaus Genuit, André Fiebig, José Coelho, Luigi Maffei, and Lisa Lavia: 'Ten questions on the soundscapes of the built environment'. *Building and Environment* (Aug. 2016), vol. 108 (cit. on p. 155).

[Kan21] Kang, Namkyoo, Young June Sah, and Sangwon Lee: 'Effects of visual and auditory cues on haptic illusions for active and passive touches in mixed reality'. *International Journal of Human-Computer Studies* (2021), vol. 150: p. 102613 (cit. on p. 27).

[Kan18] Kang, Xiaomin, Baoqi Huang, and Guodong Qi: 'A Novel Walking Detection and Step Counting Algorithm Using Unconstrained Smartphones'. *Sensors* (2018), vol. 18(1) (cit. on p. 159).

[Kao15] Kao, Hsin-Liu (Cindy), Artem Dementyev, Joseph A. Paradiso, and Chris Schmandt: 'NailO: Fingernails as an Input Surface'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: Association for Computing Machinery, 2015: pp. 3015–3018 (cit. on p. 72).

[Kar20] Karakaya, Ahmet-Serdar, Jonathan Hasenburg, and David Bermbach: 'SimRa: Using crowdsourcing to identify near miss hotspots in bicycle traffic'. *Pervasive and Mobile Computing* (2020), vol. 67: p. 101197 (cit. on p. 94).

[Kas16] Kasemsadeh, Ben and Luke LaPointe: 'Inductive Sensing Touch-On-Metal Buttons Design Guide'. *Application Report SNOA951* (2016), vol. (cit. on p. 20).

[Kat99] Kato, Hirokazu and Mark Billinghurst: 'Marker tracking and hmd calibration for a video-based augmented reality conferencing system'. *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. IEEE. 1999: pp. 85–94 (cit. on p. 24).

[Kau21] Kaul, Oliver Beren and Domin, Andreas and Rohs, Michael and Simon, Benjamin and Schrapel, Maximilian: 'VRTactileDraw: A Virtual Reality Tactile Pattern Designer for Complex Spatial Arrangements of Actuators'. *IFIP Conference on Human-Computer Interaction*. Springer. 2021: pp. 212–233 (cit. on p. 267).

[Kaw18] Kawsar, Fahim, Chulhong Min, Akhil Mathur, Alessandro Montanari, Utku Günay Acer, and Marc Van den Broeck: 'esense: Open earable platform for human sensing'. *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*. 2018: pp. 371–372 (cit. on p. 156).

[Kay77] Kay, Alan and Adele Goldberg: 'Personal dynamic media'. *Computer* (1977), vol. 10(3): pp. 31–41 (cit. on p. 17).

[Kee20] Keef, Colin V, Laure V Kayser, Stazia Tronboll, Cody W Carpenter, Nicholas B Root, Mickey Finn III, Timothy F O'Connor, Sami N Abuhamdieh, Daniel M Davies, Rory Runser, et al.: 'Virtual Texture Generated Using Elastomeric Conductive Block Copolymer in a Wireless Multimodal Haptic Glove'. *Advanced Intelligent Systems* (2020), vol. 2(4): p. 2000018 (cit. on p. 27).

[Kel06] Kela, Juha, Panu Korpipää, Jani Mäntyjärvi, Sanna Kallio, Giuseppe Savino, Luca Jozzo, and Di Marca: 'Accelerometer-Based Gesture Control for a Design Environment'. *Personal Ubiquitous Comput.* (2006), vol. 10(5): pp. 285–299 (cit. on p. 58).

[Ker12] Kerkez, Branko, Steven D. Glaser, Roger C. Bales, and Matthew W. Meadows: 'Design and performance of a wireless sensor network for catchment-scale snow and soil moisture measurements'. *Water Resources Research* (2012), vol. 48(9) (cit. on p. 95).

[Ker20] Kern, Angelika C. and Wolfgang Ellermeier: 'Audio in VR: Effects of a Soundscape and Movement-Triggered Step Sounds on Presence'. *Frontiers in Robotics and AI* (2020), vol. 7: p. 20 (cit. on pp. 13, 27, 156).

[Ket13] Ketabdar, Hamed, Amin Haji Abolhassani, and M. Roshandel: 'MagiThings: Gestural Interaction with Mobile Devices Based on Using Embedded Compass (Magnetic Field) Sensor'. *Int. J. Mob. Hum. Comput. Interact.* (2013), vol. 5: pp. 23–41 (cit. on p. 72).

[Ket12] Ketabdar, Hamed, Peyman Moghadam, Babak Naderi, and Mehran Roshandel: 'Magnetic signatures in air for mobile devices'. *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion.* 2012: pp. 185–188 (cit. on p. 56).

[Ket10a] Ketabdar, Hamed, Mehran Roshandel, and Kamer Ali Yüksel: 'Towards Using Embedded Magnetic Field Sensor for Around Mobile Device 3D Interaction'. *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services.* MobileHCI '10. Lisbon, Portugal: ACM, 2010: pp. 153–156 (cit. on p. 72).

[Ket10b] Ketabdar, Hamed, Kamel A Yüksel, Amirhossein Jahnbekam, Mehram Roshandel, and Daria Skripko: 'MagiSign: User Identification/Authentication'. *Proc. Of UBICOMM'10* (2010), vol. (cit. on p. 56).

[Ket10c] Ketabdar, Hamed, Kamer Ali Yüksel, and Mehran Roshandel: 'MagiTact: interaction with mobile devices based on compass (magnetic) sensor'. Feb. 2010: pp. 413–414 (cit. on p. 72).

[Kie13] Kienzle, Wolf and Ken Hinckley: 'Writing Handwritten Messages on a Small Touchscreen'. *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services.* MobileHCI '13. Munich, Germany: ACM, 2013: pp. 179–182 (cit. on p. 70).

[Kig11] Kight, Caitlin R. and J. Swaddle: 'How and why environmental noise impacts animals: an integrative, mechanistic review.' *Ecology letters* (2011), vol. 14 10: pp. 1052–61 (cit. on p. 154).

[Kir09] Kirby, Richard: 'Development of a real-time performance measurement and feedback system for alpine skiers'. *Sports Technology* (2009), vol. 2(1-2): pp. 43–52 (cit. on p. 95).

[Kir12] Kirk, David, Shahram Izadi, Otmar Hilliges, Richard Banks, Stuart Taylor, and Abigail Sellen: 'At home with surface computing?' *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2012: pp. 159–168 (cit. on p. 3).

[Kis10] Kise, Koichi, Megumi Chikano, Kazumasa Iwata, Masakazu Iwamura, Seiichi Uchida, and Shinichiro Omachi: 'Expansion of Queries and Databases for Improving the Retrieval Accuracy of Document Portions: An Application to a Camera-pen System'. *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems.* DAS '10. Boston, Massachusetts, USA: ACM, 2010: pp. 309–316 (cit. on p. 33).

[Kni14] Knibbe, Jarrod, Diego Martinez Plasencia, Christopher Bainbridge, Chee-Kin Chan, Jiawei Wu, Thomas Cable, Hassan Munir, and David Coyle: 'Extending Interaction for Smart Watches: Enabling Bimanual Around Device Control'. *CHI '14 Extended Abstracts on Human Factors in Computing Systems.* CHI EA '14. Toronto, Ontario, Canada: ACM, 2014: pp. 1891–1896 (cit. on pp. 25, 72).

[Kob08] Kobayashi, Hiroki, Ryoko Ueoka, and Michitaka Hirose: 'Wearable Forest-Feeling of Belonging to Nature'. *Proceedings of the 16th ACM International Conference on Multimedia.* MM '08. Vancouver, British Columbia, Canada: Association for Computing Machinery, 2008: pp. 1133–1134 (cit. on p. 155).

[Kom15] Komninos, Andreas, Mark D Dunlop, David Rowe, Allan Hewitt, and Steven Coull: 'Using degraded music quality to encourage a health improving walking pace: BeatClearWalker'. *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. IEEE. 2015: pp. 57–64 (cit. on p. 155).

[Kop11] Koptyug, Andrey and Leonid Kuzmin: 'Experimental field studies of the cross-country ski running surface interaction with snow'. *Procedia Engineering* (2011), vol. 13. 5th Asia-Pacific Congress on Sports Technology (APCST): pp. 23–29 (cit. on p. 95).

[Kra99] Kramer, Gregory et al.: 'The sonification report: Status of the field and research agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display'. (Jan. 1999), vol. (cit. on p. 155).

[Kri09] Krishnan, Narayanan C., Colin Juillard, Dirk Colbry, and Sethuraman Panchanathan: 'Recognition of Hand Movements Using Wearable Accelerometers'. *J. Ambient Intell. Smart Environ.* (2009), vol. 1(2): pp. 143–155 (cit. on pp. 33, 39).

[Kur21] Kurowski, Mariusz, Andrzej Sroczyński, Georgis Bogdanis, and Andrzej Czyżewski: 'An Automated Method for Biometric Handwritten Signature Authentication Employing Neural Networks'. *Electronics* (2021), vol. 10(4): p. 456 (cit. on p. 56).

[Kyu08] Kyung, Ki-Uk and Jun-Young Lee: 'WUbi-Pen: Windows Graphical User Interface Interacting with Haptic Feedback Stylus'. *ACM SIGGRAPH 2008 New Tech Demos*. SIGGRAPH '08. Los Angeles, California: Association for Computing Machinery, 2008 (cit. on p. 56).

[lEc95] l'Eclairage, Commission Internationale de: 'Method of Measuring and Specifying Colour Rendering Properties of Light Sources, CIE Publication 13.3-1995, Vienna, Austria, 1995,20 pp'. *Color Research & Application* (1995), vol. 20(3) (cit. on p. 118).

[Lai17] Lai, Songxuan, Lianwen Jin, and Weixin Yang: 'Online signature verification using recurrent neural network and length-normalized path signature descriptor'. *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*. Vol. 1. IEEE. 2017: pp. 400–405 (cit. on p. 55).

[Lam89] Lam, Chan F. and David Kamins: 'Signature recognition through spectral analysis'. *Pattern Recognition* (1989), vol. 22(1): pp. 39–44 (cit. on pp. 59, 64).

[Lan07] LANDY, MICHAEL S: 'A gloss on surface properties'. *Nature* (2007), vol. 447(7141): pp. 158–159 (cit. on p. 11).

[Law20] LAWTON, MARK, STUART CUNNINGHAM, and IAN CONVERY: 'Nature soundscapes: an audio augmented reality experience'. *Proceedings of the 15th International Conference on Audio Mostly*. 2020: pp. 85–92 (cit. on pp. 155, 156).

[Lax19] LAXMI, V., JAYDIP DEY, KOMAL KALAWAPUDI, R. VIJAY, and RAKESH KUMAR: 'An innovative approach of urban noise monitoring using cycle in Nagpur, India'. *Environmental Science and Pollution Research* (2019), vol. 26: pp. 36812–36819 (cit. on pp. 92, 94).

[Led86] LEDERMAN, SUSAN, GEORGIE THORNE, and BILL JONES: 'Perception of Texture by Vision and Touch. Multidimensionality and Intersensory Integration'. *Journal of experimental psychology. Human perception and performance* (June 1986), vol. 12: pp. 169–80 (cit. on p. 15).

[Led09] LEDERMAN, SUSAN J and ROBERTA L KLATZKY: 'Haptic perception: A tutorial'. *Attention, Perception, & Psychophysics* (2009), vol. 71(7): pp. 1439–1459 (cit. on p. 15).

[Lee08] LEE, DAH-JYE, YUCHOU CHANG, JAMES K. ARCHIBALD, and CLINT PITZAK: 'Matching book-spine images for library shelf-reading process automation'. *2008 IEEE International Conference on Automation Science and Engineering* (2008), vol.: pp. 738–743 (cit. on p. 134).

[Lee04] LEE, JOHNNY C., PAUL H. DIETZ, DARREN LEIGH, WILLIAM S. YERAZUNIS, and SCOTT E. HUDSON: 'Haptic Pen: A Tactile Feedback Stylus for Touch Screens'. UIST '04. Santa Fe, NM, USA: Association for Computing Machinery, 2004: pp. 291–294 (cit. on p. 56).

[Lee03] LEE, MINKYUNG and WOONTACK WOO: 'ARKB: 3D vision-based Augmented Reality Keyboard'. *ICAT*. 2003 (cit. on p. 26).

[Lee22] LEE, SHIH-HSIUNG and HSUAN-CHIH KU: 'SIAR: Signing in the Air Using Finger Tracking Technology for Authentication on Embedded System'. *2022 IEEE International Conference on Consumer Electronics (ICCE)*. 2022: pp. 1–2 (cit. on p. 55).

[Lee07] LEE, TAEHEE and TOBIAS HOLLERER: 'Handy AR: Markerless inspection of augmented reality objects using fingertip tracking'. *2007 11th IEEE International Symposium on Wearable Computers*. IEEE. 2007: pp. 83–90 (cit. on p. 26).

[Lei15] Leiva, Luis A., Alireza Sahami, Alejandro Catala, Niels Henze, and Albrecht Schmidt: 'Text Entry on Tiny QWERTY Soft Keyboards'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, 2015: pp. 669–678 (cit. on pp. 70, 72).

[Les08] Lesaki, Akiko, Akihiro Somada, Asako Kimura, Fumihisa Shibata, and Hideyuki Tamura: 'Psychophysical Influence on Tactual Impression by Mixed-Reality Visual Stimulation'. *2008 IEEE Virtual Reality Conference*. 2008: pp. 265–266 (cit. on p. 27).

[Lev66] Levenshtein, Vladimir Iosifovich: 'Binary codes capable of correcting deletions, insertions and reversals'. *Soviet Physics Doklady* (1966), vol. 10(8): pp. 707–710 (cit. on p. 136).

[Lev16] Levesque, Vincent, Juan Manuel Cruz-Hernandez, Amaya Weddle, and David M Birnbaum: *System and method for simulated physical interactions with haptic effects*. US Patent 9,330,544. May 2016 (cit. on p. 23).

[Lev11] Levesque, Vincent, Louise Oram, Karon MacLean, Andy Cockburn, Nicholas D Marchuk, Dan Johnson, J Edward Colgate, and Michael A Peshkin: 'Enhancing physicality in touch interaction with programmable friction'. *Proceedings of the SIGCHI conference on human factors in computing systems*. 2011: pp. 2481–2490 (cit. on pp. 22, 23).

[Lev18] Levy, Alona, Ben Nassi, Yuval Elovici, and Erez Shmueli: 'Handwritten signature verification using wrist-worn devices'. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2018), vol. 2(3): pp. 1–26 (cit. on pp. 55, 60).

[Lew16] Lewis, Antwane, Yanyan Li, and Mengjun Xie: 'Real time motion-based authentication for smartwatch'. *2016 IEEE Conference on Communications and Network Security (CNS)*. 2016: pp. 380–381 (cit. on p. 55).

[Ley19] Ley-Flores, Judith, Frédéric Bevilacqua, Nadia Bianchi-Berthouze, and Ana Taiadura-Jiménez: 'Altering body perception and emotion in physically inactive people through movement sonification'. *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 2019: pp. 1–7 (cit. on pp. 13, 156).

[Li04] Li, F. F.: 'Handwriting authentication by envelopes of sound signature'. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004*. Vol. 1. 2004: 401–404 Vol.1 (cit. on pp. 34, 38, 39, 56, 64).

[Li20] Li, Gen and Hiroyuki Sato: 'Handwritten Signature Authentication Using Smartwatch Motion Sensors'. *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE. 2020: pp. 1589–1596 (cit. on p. 56).

[Li21] Li, Gen, Lingfeng Zhang, and Hiroyuki Sato: 'In-air Signature Authentication Using Smartwatch Motion Sensors'. *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE. 2021: pp. 386–395 (cit. on p. 55).

[Li19] Li, Oyun-Erdene Enkhjargal1and Kai Way: 'Subjective ratings of floor slippery on common indoor and outdoor floors'. *International Journal of Engineering and Technology* (2019), vol. 11(4) (cit. on p. 2).

[Li11] Li, Wenzhe and Tracy Anne Hammond: 'Recognizing Text Through Sound Alone'. *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011*. 2011 (cit. on p. 34).

[Lig14] Light, Ann and Clodagh Miskelly: *Design for Sharing*. Tech. rep. Northumbria University, Nov. 2014 (cit. on pp. 131, 132).

[Lim19] Lim, Jongho, Yongjae Yoo, Hanseul Cho, and Seungmoon Choi: 'Touchphoto: Enabling independent picture taking and understanding for visually-impaired users'. *2019 International Conference on Multimodal Interaction*. 2019: pp. 124–134 (cit. on p. 23).

[Liu03] Liu, Cheng-Lin, Kazuki Nakashima, Hiroshi Sako, and Hiromichi Fujisawa: 'Handwritten digit recognition: Benchmarking of state-of-the-art techniques'. (Oct. 2003), vol. 36: pp. 2271–2285 (cit. on p. 33).

[LLC19] LLC, Google: *ARCore - Google Developers Documentation*. `https://developers.google.com/ar`. 2019 (cit. on pp. 135, 140).

[Löc10] Löchtefeld, Markus, Sven Gehring, Johannes Schöning, and Antonio Krüger: 'ShelfTorchlight: Augmenting a Shelf using a Camera Projector Unit'. *Adjunct Proceedings of the Eighth International Conference on Pervasive Computing*. Jan. 2010: pp. 1–4 (cit. on p. 134).

[Lon00] Long, Allan, James Landay, Lawrence Rowe, and Joseph Michiels: 'Visual similarity of pen gestures'. Jan. 2000: pp. 360–367 (cit. on p. 58).

[Lop11] Lopes, Pedro, Ricardo Jota, and Joaquim A Jorge: 'Augmenting touch interaction through acoustic sensing'. *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. 2011: pp. 53–56 (cit. on p. 19).

[Lou84] Louie, J.K., C.Y. Kuo, M.D. Gutierrez, and C.D. Mote: 'Surface EMG and torsion measurements during snow skiing: Laboratory and field tests'. *Journal of Biomechanics* (1984), vol. 17(10): pp. 713–724 (cit. on p. 95).

[Lu21] Lu, Xueshi, Difeng Yu, Hai-Ning Liang, and Jorge Goncalves: 'iText: Hands-free text entry on an imaginary keyboard for augmented reality systems'. *The 34th Annual ACM Symposium on User Interface Software and Technology*. 2021: pp. 815–825 (cit. on p. 26).

[Lyo20] Lyons, Kent: 'Wearable Magnetic Field Sensing for Finger Tracking'. ISWC '20. Virtual Event, Mexico: Association for Computing Machinery, 2020: pp. 63–67 (cit. on p. 72).

[Mac12] MacKenzie, I Scott: 'Human-computer interaction: An empirical research perspective'. (2012), vol. (cit. on pp. 5, 9, 13, 16, 17, 23).

[Mad11] Madgwick, S. O. H., A. J. L. Harrison, and R. Vaidyanathan: 'Estimation of IMU and MARG orientation using a gradient descent algorithm'. *2011 IEEE International Conference on Rehabilitation Robotics*. 2011: pp. 1–7 (cit. on p. 78).

[Mae05] Maeda, T., H. Ando, T. Amemiya, N. Nagaya, M. Sugimoto, and M. Inami: 'Shaking the World: Galvanic Vestibular Stimulation as a Novel Sensation Interface'. *ACM SIGGRAPH 2005 Emerging Technologies*. SIGGRAPH '05. Los Angeles, California: Association for Computing Machinery, 2005: 17–es (cit. on p. 13).

[Mal19] Malek, S. A., S. D. Glaser, and R. C. Bales: 'Wireless Sensor Networks for Improved Snow Water Equivalent and Runoff Estimates'. *IEEE Access* (2019), vol. 7: pp. 18420–18436 (cit. on p. 95).

[Mal53] Mallinckrodt, Edward, AL Hughes, and William Sleator Jr: 'Perception by the skin of electrically induced vibrations'. *Science* (1953), vol. 118(3062): pp. 277–278 (cit. on p. 22).

[Mao19] Mao, Ruiquan, Manuel Lagunas, Belen Masia, and Diego Gutierrez: 'The effect of motion on the perception of material appearance'. *ACM Symposium on Applied Perception 2019*. 2019: pp. 1–9 (cit. on p. 11).

[Mar13] Mariette, Nicholas: 'Human factors research in audio augmented reality'. *Human factors in augmented reality environments*. Springer, 2013: pp. 11–32 (cit. on p. 155).

[Mar18] Marquis, Raymond, Williams David Mazzella, and Tacha Hicks: 'X-marks: Too simple to be useful?' *Nowa kodyfikacja prawa karnego* (2018), vol. 49: pp. 103–110 (cit. on pp. 54, 58, 65).

[Mat11] Matsushita, Kazuhiro, Daisuke Iwai, and Kosuke Sato: 'Interactive bookshelf surface for in situ book searching and storing support'. Jan. 2011: p. 2 (cit. on p. 134).

[Mat13] Mattern, Friedemann: *Total vernetzt: Szenarien einer informatisierten Welt*. Springer-Verlag, 2013 (cit. on p. 17).

[Mat01] Mattern, Friedemann: 'Ubiquitous computing'. *Internet at Future, Jahrbuch Telekommunikation und Gesellschaft* (2001), vol.: pp. 52–61 (cit. on p. 16).

[Max18a] Maximilian Schrapel, Luise aus der Fünten (EZN): *Pentelligence: Handschrifterkennung mittels Audio und Bewegungsdaten von Stiften*. DE Patent 10 2018 107 409 A1. 2018 (cit. on pp. 31, 269).

[Max18b] Maximilian Schrapel & Anne Finger & Jochen Meyer & Michael Rohs & Johannes Schöning & Alexandra Voit: *Ubiactivity'18: International Workshop on Integrating Physical Activity and Health Aspects in Everyday Mobility. UbiComp'18, Singapore*. 2018 (cit. on p. 268).

[Max18c] Maximilian Schrapel, Anne Finger and Michael Rohs: 'Integrating Recommended Physical Activity in Everyday Mobility'. *Accepted workshop papers at the workshop on Augmented Humanity using Wearable and Mobile Devices for Health and Wellbeing at MobileHCI'18* (2018), vol. (cit. on p. 268).

[McC15] McCarthy, Orla, Brian Caulfield, and Gerard Deenihan: 'Evaluating the quality of inter-urban cycleways'. *Case Studies on Transport Policy* (Nov. 2015), vol. (cit. on p. 93).

[McG20] McGill, Mark, Stephen Brewster, David McGookin, and Graham Wilson: 'Acoustic transparency and the changing soundscape of auditory mixed reality'. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020: pp. 1–16 (cit. on pp. 156, 157, 174, 175).

[McG11] McGookin, David and Stephen Brewster: 'PULSE: An Auditory Display to Provide a Social Vibe'. *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*. IwS '11. Stockholm, Sweden: Association for Computing Machinery, 2011: pp. 12–15 (cit. on p. 155).

[McG09] McGookin, David K., Stephen A. Brewster, and Pablo Priego: 'Audio Bubbles: Employing Non-speech Audio to Support Tourist Wayfinding'. *HAID*. 2009 (cit. on pp. 155–157).

[McI19] McIntosh, Jess, Paul Strohmeier, Jarrod Knibbe, Sebastian Boring, and Kasper Hornbæk: 'Magnetips: Combining Fingertip Tracking and Haptic Feedback for Around-Device Interaction'. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: ACM, 2019: 408:1–408:12 (cit. on pp. 25, 72, 73).

[Mee20] Meena, Leetesh, Vijay Kumar Chaurasiya, Neetesh Purohit, and Dhananjay Singh: 'Comparison of SVM and Random Forest Methods for Online Signature Verification'. *International Conference on Intelligent Human Computer Interaction*. Springer. 2020: pp. 288–299 (cit. on p. 55).

[mei20] mein-dienstrad.de: *Ecosense: Erfassung und Analyse von Fahrradverkehrsdaten (accessed September 1, 2020)*. 2020 (cit. on pp. 94, 98).

[Men10] Menzer, Fritz, Anna Brooks, Pär Halje, Christof Faller, Martin Vetterli, and Olaf Blanke: 'Feeling in control of your footsteps: conscious gait monitoring and the auditory consequences of footsteps'. *Cognitive neuroscience* (2010), vol. 1(3): pp. 184–192 (cit. on pp. 13, 156, 157).

[Mes16] Messenger, Jon C and Lutz Gschwind: 'Three generations of Telework: New ICT s and the (R) evolution from Home Office to Virtual Office'. *New Technology, Work and Employment* (2016), vol. 31(3): pp. 195–208 (cit. on p. 124).

[Mia20] Miah, S., E. Milonidis, I. Kaparias, and N. Karcanias: 'An Innovative Multi-Sensor Fusion Algorithm to Enhance Positioning Accuracy of an Instrumented Bicycle'. *IEEE Transactions on Intelligent Transportation Systems* (2020), vol. 21(3) (cit. on p. 94).

[Mic19] Microsoft: *Bing Visual Search*. https://www.bing.com/visualsearch. 2019 (cit. on p. 134).

[Mil94] Milgram, Paul, Haruo Takemura, Akira Utsumi, and Fumio Kishino: 'Augmented reality: A class of displays on the reality-virtuality continuum'. *Telemanipulator and Telepresence Technologies* (Jan. 1994), vol. 2351 (cit. on p. 24).

[Mit19] Mitake, Hiroto, Hiroki Watanabe, and Masanori Sugimoto: 'Footsteps and Inertial Data-Based Road Surface Condition Recognition Method'. *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. MUM '19. Pisa, Italy: Association for Computing Machinery, 2019 (cit. on pp. 94, 100, 184).

[Miu07] Miura, M., S. Kunifuji, and Y. Sakamoto: 'AirTransNote: An Instant Note Sharing and Reproducing System to Support Students Learning'. *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*. 2007: pp. 175–179 (cit. on p. 32).

[MIY02] MIYAGAWA, Tohru, Yoshimichi YONEZAWA, Kazunori Itoh, and Masami Hashimoto: 'Handwritten Pattern Reproduction Using 3D Inertial Measurement of Handwriting Movement'. (Jan. 2002), vol. 38: pp. 1–8 (cit. on p. 33).

[Mol13] Molesworth, Brett R.C., Marion Burgess, and Daniel Kwon: 'The use of noise cancelling headphones to improve concurrent task performance in a noisy environment'. *Applied Acoustics* (2013), vol. 74(1): pp. 110–115 (cit. on pp. 154, 156).

[Mon22] Monica, Riccardo and Jacopo Aleotti: 'Evaluation of the Oculus Rift S tracking system in room scale virtual reality'. *Virtual Reality* (2022), vol.: pp. 1–11 (cit. on p. 25).

[Moo17] Moon, Han-Cheol, Se-In Jang, Kangrok Oh, and Kar-Ann Toh: 'An in-air signature verification system using Wi-Fi signals'. *Proceedings of the 2017 4th International Conference on Biomedical and Bioinformatics Engineering*. 2017: pp. 133–138 (cit. on p. 55).

[Moo65] Moore, Gordon E et al.: *Cramming more components onto integrated circuits*. 1965 (cit. on p. 1).

[Mor83] Morley, JW, AW Goodwin, and I Darian-Smith: 'Tactile discrimination of gratings'. *Experimental brain research* (1983), vol. 49(2): pp. 291–299 (cit. on p. 15).

[Mot08] Motamedi, Nima: 'Hd touch: multi-touch and object sensing on a high definition lcd tv'. *CHI'08 Extended Abstracts on Human Factors in Computing Systems*. 2008: pp. 3069–3074 (cit. on p. 18).

[Mot07] Motoyoshi, Isamu, Shin'ya Nishida, Lavanya Sharan, and Edward H Adelson: 'Image statistics and the perception of surface qualities'. *Nature* (2007), vol. 447(7141): pp. 206–209 (cit. on p. 11).

[Mul14] Mullenbach, Joe, Craig Shultz, J Edward Colgate, and Anne Marie Piper: 'Exploring affective communication through variable-friction surface haptics'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014: pp. 3963–3972 (cit. on p. 23).

[Mül20] Müller, Benjamin, Andreas Liebl, and Noemi Martin: 'Is wearing active-noise-cancelling headphones in open space offices beneficial for cognitive performance and employee satisfaction? Results of two laboratory studies.' Aug. 2020 (cit. on pp. 154, 156).

[Mül94] Müller, Erich: 'Analysis of the biomechanical characteristics of different swinging techniques in alpine skiing'. *Journal of sports sciences* (July 1994), vol. 12: pp. 261–78 (cit. on pp. 95, 99).

[Mun02] Munich, Mario E. and Pietro Perona: 'Visual Input for Pen-Based Computers'. *IEEE Trans. Pattern Anal. Mach. Intell.* (2002), vol. 24(3): pp. 313–328 (cit. on p. 33).

[Mün18] Münzel, Thomas, Mette Sørensen, Frank Schmidt, Erwin Schmidt, Sebastian Steven, Swenja Kröller-Schön, and Andreas Daiber: 'The adverse effects of environmental noise exposure on oxidative stress and cardiovascular risk'. *Antioxidants & redox signaling* (2018), vol. 28(9): pp. 873–908 (cit. on p. 154).

[Mur03] Muramatsu, Daigo and Takashi Matsumoto: 'An HMM online signature verifier incorporating signature trajectories'. *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* IEEE. 2003: pp. 438–442 (cit. on p. 55).

[Mur07] Muramatsu, Daigo and Takashi Matsumoto: 'Effectiveness of pen pressure, azimuth, and altitude features for online signature verification'. *International Conference on Biometrics*. Springer. 2007: pp. 503–512 (cit. on pp. 56, 64).

[Mur12a] Muramatsu, Yuichi, Mihoko Niitsuma, and Trygve Thomessen: 'Perception of tactile sensation using vibrotactile glove interface'. *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE. 2012: pp. 621–626 (cit. on p. 27).

[Mur12b] Murugappan, Sundar, Niklas Elmqvist, and Karthik Ramani: 'Extended multitouch: recovering touch posture and differentiating users using a depth camera'. *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 2012: pp. 487–496 (cit. on p. 20).

[Nab95] Nabeshima, Shinji, Shinichirou Yamamoto, Kiyoshi Agusa, and Toshio Taguchi: 'MEMO-PEN: A New Input Device'. *Conference Companion on Human Factors in Computing Systems*. CHI '95. Denver, Colorado, USA: ACM, 1995: pp. 256–257 (cit. on p. 33).

[Nag18] Nagel, Lukas: 'Using pen movement and audio data to verify signatures'. Bachelor Thesis. Leibniz Universität Hannover, 2018 (cit. on pp. k, 53).

[Nan16] Nandakumar, Rajalakshmi, Vikram Iyer, Desney Tan, and Shyamnath Gollakota: 'FingerIO: Using Active Sonar for Fine-Grained Finger Tracking'. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2016: pp. 1515–1525 (cit. on p. 72).

[Nev15] Nevetha, M. P. and A. Baskar: 'Automatic Book Spine Extraction and Recognition for Library Inventory Management'. *WCI*. 2015 (cit. on p. 134).

[New79] Newman, William M. and Robert F. Sproull, eds.: *Principles of Interactive Computer Graphics (2Nd Ed.)* New York, NY, USA: McGraw-Hill, Inc., 1979 (cit. on p. 137).

[Ngo15] Ngoc Diep, Nguyen, Cuong Pham, and Tu Minh Phuong: 'SigVer3D: Accelerometer based verification of 3-D signatures on mobile devices'. *Knowledge and Systems Engineering*. Springer, 2015: pp. 353–365 (cit. on p. 55).

[Ngu15] Nguyen, Anh and Amy Banic: *3DTouch: A wearable 3D input device for 3D applications*. IEEE, 2015 (cit. on p. 26).

[Nif18] Niforatos, E., A. Fedosov, M. Langheinrich, and I. Elhart: 'Augmenting Humans on the Slope: Two Electronic Devices That Enhance Safety and Decision Making'. *IEEE Consumer Electronics Magazine* (2018), vol. 7(3): pp. 81–89 (cit. on p. 94).

[Nif16] Niforatos, Evangelos, Ivan Elhart, Anton Fedosov, and Marc Langheinrich: 's-Helmet: A Ski Helmet for Augmenting Peripheral Perception'. Feb. 2016 (cit. on p. 94).

[Nif17] Niforatos, Evangelos, Anton Fedosov, Ivan Elhart, and Marc Langheinrich: 'Augmenting skiers' peripheral perception'. Sept. 2017: pp. 114–121 (cit. on p. 94).

[Nis04] Nistér, David, Oleg Naroditsky, and James Bergen: 'Visual odometry'. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 1. Ieee. 2004: pp. I–I (cit. on p. 25).

[Nor10] Nordahl, Rolf, Amir Berrezag, Smilen Dimitrov, Luca Turchet, Vincent Hayward, and Stefania Serafin: 'Preliminary experiment combining virtual reality haptic shoes and audio synthesis'. *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications.* Springer. 2010: pp. 123–129 (cit. on pp. 27, 28).

[ONe99] O'Neill, Daniel Fulham and Mark R. McGlone: 'Injury Risk in First-Time Snowboarders Versus First-Time Skiers'. *The American Journal of Sports Medicine* (1999), vol. 27(1). PMID: 9934425: pp. 94–97 (cit. on p. 95).

[ODo86] O'Donnell, Francis XD and Fred W Billmeyer Jr: 'Psychometric scaling of gloss'. *Review and evaluation of appearance: Methods and techniques. ASTM STP* (1986), vol. 914: pp. 14–32 (cit. on p. 11).

[Oak14] Oakley, Ian and Doyoung Lee: 'Interaction on the Edge: Offset Sensing for Small Devices'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI '14. Toronto, Ontario, Canada: ACM, 2014: pp. 169–178 (cit. on p. 72).

[Oga15] Ogata, Masa and Michita Imai: 'SkinWatch: skin gesture interaction for smart watch'. *AH '15.* 2015 (cit. on pp. 25, 72).

[Oga12] Ogata, Masa, Yuta Sugiura, Hirotaka Osawa, and Michita Imai: 'IRing: Intelligent Ring Using Infrared Reflection'. *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology.* UIST '12. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2012: pp. 131–136 (cit. on p. 72).

[Oh19] Oh, Seungjae, Gyeore Yun, Chaeyong Park, Jinsoo Kim, and Seungmoon Choi: 'VibEye: Vibration-Mediated Object Recognition for Tangible Interactive Applications'. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019: pp. 1–12 (cit. on p. 118).

[Ohn19] Ohnishi, Ayumi, Isao Nishiyama, Tsutomu Terada, and Masahiko Tsukamoto: 'An Auditory Feedback System to Improve the Foot Pressure Balance for Runners'. *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia*. MoMM2019. Munich, Germany: Association for Computing Machinery, 2019: pp. 94–101 (cit. on pp. 13, 156).

[Oka12] Okamoto, Shogo, Hikaru Nagano, and Yoji Yamada: 'Psychophysical dimensions of tactile perception of textures'. *IEEE Transactions on Haptics* (2012), vol. 6(1): pp. 81–93 (cit. on p. 15).

[Oli06] Oliver, Nuria and Fernando Flores-Mangas: 'MPTrain: a mobile, music and physiology-based personal trainer'. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*. 2006: pp. 21–28 (cit. on p. 155).

[One13] Oney, Stephen, Chris Harrison, Amy Ogan, and Jason Wiese: 'ZoomBoard: A Diminutive Qwerty Soft Keyboard Using Iterative Zooming for Ultra-small Devices'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '13. Paris, France: ACM, 2013: pp. 2799–2802 (cit. on p. 71).

[Ope20] OpenCycleMap: *OpenCycleMap - the OpenStreetMap Cycle Map (accessed September 1, 2020)*. 2020 (cit. on pp. 94, 111).

[Org11] Organization, World Health et al.: *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe, 2011 (cit. on pp. 154, 155).

[Org18] Organization, World Health et al.: 'Environmental noise guidelines for the European region'. (2018), vol. (cit. on p. 154).

[Org10] Organization, World Health: *Global Recommendations on Physical Activity for Health*. World Health Organization, 2010 (cit. on pp. 155, 157).

[Osb80] Osborne, A.: *An Introduction to Microcomputers: Basic concepts*. An Introduction to Microcomputers. Osborne/McGraw-Hill, 1980 (cit. on p. 37).

[Ozi17] Ozioko, Oliver, William Taube, Marion Hersh, and Ravinder Dahiya: 'SmartFingerBraille: A tactile sensing and actuation based communication glove for deafblind people'. *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*. IEEE. 2017: pp. 2014–2018 (cit. on p. 27).

[P H00] P. H. Halldin, M. Aare and H. von Hoist: 'An Experimental Safety Helmet with Improved Rotational Protection'. *Crash Prevention and Injury Control* (Jan. 2000), vol. 2 (cit. on p. 94).

[Pan18] Pan, Tse-Yu, Chih-Hsuan Kuo, Hou-Tim Liu, and Min-Chun Hu: 'Handwriting trajectory reconstruction using low-cost imu'. *IEEE Transactions on Emerging Topics in Computational Intelligence* (2018), vol. 3(3): pp. 261–270 (cit. on p. 56).

[Par00] Paradiso, J. A., K. Hsiao, J. Strickon, J. Lifton, and A. Adler: 'Sensor systems for interactive surfaces'. *IBM Systems Journal* (2000), vol. 39(3.4): pp. 892–914 (cit. on p. 18).

[Par90] Parizeau, Marc and Réjean Plamondon: 'A Comparative Analysis of Regional Correlation, Dynamic Time Warping, and Skeletal Tree Matching for Signature Verification'. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (Aug. 1990), vol. 12: pp. 710–717 (cit. on p. 55).

[Par19a] Parizi, Farshid Salemi, Eric Whitmire, and Shwetak N. Patel: 'AuraRing: Precise Electromagnetic Finger Tracking'. *IMWUT* (2019), vol. 3: 150:1–150:28 (cit. on pp. 25, 72).

[Par11] Park, Gunhyuk, Seungmoon Choi, Kyunghun Hwang, Sunwook Kim, Jaecheon Sa, and Moonchae Joung: 'Tactile effect design and evaluation for virtual buttons on a mobile device touchscreen'. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. 2011: pp. 11–20 (cit. on p. 22).

[Par20] Park, Keunwoo, Daehwa Kim, Seongkook Heo, and Geehyuk Lee: 'MagTouch: Robust Finger Identification for a Smartwatch Using a Magnet Ring and a Built-in Magnetometer'. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, 2020: pp. 1–13 (cit. on pp. 71, 72).

[Par19b] Park, Keunwoo and Geehyuk Lee: 'FingMag: Finger Identification Method for Smartwatch'. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI EA '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019 (cit. on p. 72).

[Pec17] Pece, Fabrizio, Juan Jose Zarate, Velko Vechev, Nadine Besse, Olexandr Gudozhnik, Herbert Shea, and Otmar Hilliges: 'MagTics: Flexible and Thin Form Factor Magnetic Actuators for Dynamic and Wearable Haptic Feedback'. *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. UIST '17. Québec City, QC, Canada: Association for Computing Machinery, 2017: pp. 143–154 (cit. on p. 72).

[Ped11] Pedregosa, F. et al.: 'Scikit-learn: Machine Learning in Python'. *Journal of Machine Learning Research* (2011), vol. 12: pp. 2825–2830 (cit. on pp. 60, 120–122).

[Pel10] Pelegris, P., K. Banitsas, T. Orbach, and K. Marias: 'A novel method to detect Heart Beat Rate using a mobile phone'. *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 2010 (cit. on pp. 116, 127).

[Pen50] Penfield, Wilder and Theodore Rasmussen: 'The cerebral cortex of man; a clinical study of localization of function.' (1950), vol. (cit. on p. 9).

[Per13] Perrault, Simon, Eric Lecolinet, James Eagan, and Yves Guiard: 'WatchIt: Simple gestures and eyes-free interaction for wristwatches and bracelets'. Apr. 2013: pp. 1451–1460 (cit. on p. 72).

[Pla18] Plazak, Joseph and Marta Kersten-Oertel: 'A Survey on the Affordances of "Hearables"'. *Inventions* (2018), vol. 3(3): p. 48 (cit. on p. 156).

[Poh15a] Pohl, Henning, Markus Krause, and Michael Rohs: 'One-button recognizer: exploiting button pressing behavior for user differentiation'. Sept. 2015: pp. 403–407 (cit. on p. 56).

[Poh16] Pohl, Henning, Dennis Stanke, and Michael Rohs: 'EmojiZoom: Emoji Entry via Large Overview Maps'. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '16. Florence, Italy: ACM, 2016: pp. 510–517 (cit. on p. 71).

[Poh15b] Pohl, Henning and Becke, Dennis and Wagner, Eugen and Schrapel, Maximilian and Rohs, Michael: 'Wrist compression feedback by pneumatic actuation'. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 2015: pp. 315–318 (cit. on p. 268).

[Pra14] Prakhinkit, Susaree, Siriluck Suppapitiporn, Hirofumi Tanaka, and Daroonwan Suksom: 'Effects of Buddhism Walking Meditation on Depression, Functional Fitness, and Endothelium-Dependent Vasodilation in Depressed Elderly'. *The Journal of Alternative and Complementary Medicine* (2014), vol. 20(5). PMID: 24372522: pp. 411–416 (cit. on pp. 154, 155, 157).

[Pre73] Preiser, Wolfgang F. E.: 'Analysis of Pedestrian Velocity and Stationary Behavior in a Shopping Mall'. English. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Zuletzt aktualisiert - 2021-07-28. PhD thesis. 1973: p. 107 (cit. on pp. 2, 11, 27, 153).

[Qin18] Qin, Ryan, Suwen Zhu, Yu-Hao Lin, Yu-Jung Ko, and Xiaojun Bi: 'Optimal-T9: An Optimized T9-like Keyboard for Small Touchscreen Devices'. *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*. ISS '18. Tokyo, Japan: ACM, 2018: pp. 137–146 (cit. on p. 71).

[Quo09] Quoc, Nguyen-Huu and Won-Ho Choi: 'A framework for recognition books on bookshelves'. *ICIC 2009*. 2009 (cit. on p. 134).

[Rah22] Rahman, ABM Mohaimenur, Yetong Cao, Xinliang Wei, Pu Wang, Fan Li, and Yu Wang: 'On the Feasibility of Handwritten Signature Authentication Using PPG Sensor'. *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE. 2022: pp. 719–720 (cit. on p. 56).

[Raj16] Rajesh, Kumar, Phoha Vir V, and Raina Rahul: *Authenticating users through their arm movement patterns*. 2016 (cit. on p. 56).

[Ram21] Ramachandra, Raghavendra, Sushma Venkatesh, Kiran Raja, and Christoph Busch: 'Handwritten Signature and Text based User Verification using Smartwatch'. *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 2021: pp. 5099–5106 (cit. on p. 56).

[Reb17] Rebori, Marlene K. and Peter Burge: 'Using Geospatial Analysis to Align Little Free Library Locations with Community Literacy Needs'. *Journal of Extension* (June 217), vol. 55(3): p. 6 (cit. on p. 133).

[Rei05] Reid, Josephine, Erik Geelhoed, Richard Hull, Kirsten Cater, and Ben Clayton: 'Parallel Worlds: Immersion in Location-Based Experiences'. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2005: pp. 1733–1736 (cit. on pp. 155, 156).

[Rei17] Reinders, Christoph, Hanno Ackermann, Michael Ying Yang, and Bodo Rosenhahn: *Object Recognition from very few Training Examples for Enhancing Bicycle Maps*. 2017 (cit. on p. 94).

[Rek02] Rekimoto, Jun: 'SmartSkin: an infrastructure for freehand manipulation on interactive surfaces'. *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2002: pp. 113–120 (cit. on p. 19).

[Rek99] Rekimoto, Jun and Masanori Saitoh: 'Augmented surfaces: a spatially continuous work space for hybrid computing environments'. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 1999: pp. 378–385 (cit. on pp. 3, 18, 24).

[Ren21] Renswouw, Loes van, Jelle Neerhof, Steven Vos, Pieter van Wesemael, and Carine Lallemand: 'Sensation: Sonifying the Urban Running Experience'. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021 (cit. on pp. 28, 156, 181).

[Res20] Research, Two-Photon: *Two-Photon Research Canada Inc .CAST: The instant COVID-19 test*. https://www.twophotonresearch.com. Accessed: 2021-01-02. 2020 (cit. on p. 118).

[Rey18] Reyes, Gabriel, Jason Wu, Nikita Juneja, Maxim Goldshtein, W. Keith Edwards, Gregory D. Abowd, and Thad Starner: 'SynchroWatch: One-Handed Synchronous Smartwatch Gestures Using Correlation and Magnetic Sensing'. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (Jan. 2018), vol. 1(4): 158:1–158:26 (cit. on p. 72).

[Ric17] Riche, Yann, Nathalie Henry Riche, Ken Hinckley, Sheri Panabaker, Sarah Fuelling, and Sarah Williams: 'As We May Ink?: Learning from Everyday Analog Pen Use to Improve Digital Ink Experiences.' *CHI*. 2017: pp. 3241–3253 (cit. on p. 21).

[Rod14] Rodger, Matthew W. M., William R. Young, and Cathy M. Craig: 'Synthesis of Walking Sounds for Alleviating Gait Disturbances in Parkinson's Disease'. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2014), vol. 22: pp. 543–548 (cit. on pp. 13, 156).

[Roh05] Rohs, Michael: 'Linking physical and virtual worlds with visual markers and handheld devices'. PhD thesis. ETH Zurich, 2005 (cit. on p. 24).

[Rol01] Rolland, Jannick P, Larry D Davis, and Yohan Baillot: 'A survey of tracking technologies for virtual environments'. *Fundamentals of wearable computers and augmented reality*. CRC Press, 2001: pp. 83–128 (cit. on p. 25).

[Rom21] Romat, Hugo, Andreas Fender, Manuel Meier, and Christian Holz: 'Flashpen: a high-fidelity and high-precision multi-surface pen for virtual reality'. *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE. 2021: pp. 306–315 (cit. on p. 186).

[Ros13] Roser, Max and Hannah Ritchie: 'Technological Change (last revision 03/22)'. *Our World in Data* (2013), vol. https://ourworldindata.org/technological-change (cit. on p. 1).

[Ros16] Rossitto, Chiara, Louise Barkhuus, and Arvid Engström: 'Interweaving place and story in a location-based audio drama'. *Personal and Ubiquitous Computing* (2016), vol. 20(2): pp. 245–260 (cit. on p. 155).

[Row14] Rowe, Jonathan P, Eleni V Lobene, Bradford W Mott, and James C Lester: 'Play in the museum: Designing game-based learning environments for informal education settings.' *FDG*. 2014 (cit. on p. 3).

[Rua18] Ruan, Sherry, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay: 'Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones'. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (Jan. 2018), vol. 1(4): 159:1–159:23 (cit. on p. 136).

[Ryl20] Ryll, Steffen: 'Interaction Techniques for Magnetic Pointing Devices on Smartwatches'. Bachelor Thesis. Leibniz Universität Hannover, 2020 (cit. on pp. k, 69).

[SRa15] S.Rahmdel, Payam, Richard Comley, Daming Shi, and Siobhan McElduff: 'A Review of Hough Transform and Line Segment Detection Approaches'. *VISAPP 2015 - 10th International Conference on Computer Vision Theory and Applications; VISIGRAPP, Proceedings* (Jan. 2015), vol. 1: pp. 411–418 (cit. on p. 136).

[Sad20] Sadia, Bushra, Senem Ezgi Emgin, T Metin Sezgin, and Cagatay Basdogan: 'Data-driven vibrotactile rendering of digital buttons on touchscreens'. *International Journal of Human-Computer Studies* (2020), vol. 135: p. 102363 (cit. on p. 22).

[Sad22] Sado, Garros and Schmuhl, Julian and Schroth Peer: 'Projekt 2: Notification Management (SimpleSurfaceSocialMedia)'. Student project at course Current Topics in Human-Computer Interaction. Leibniz Universität Hannover, 2022 (cit. on pp. k, 115).

[Sae14] Sae-Bae, Napa and Nasir Memon: 'Online signature verification on mobile devices'. *IEEE transactions on information forensics and security* (2014), vol. 9(6): pp. 933–947 (cit. on p. 55).

[Sak00] Sakaguchi, Rikuo, Shino Takasu, and Tetsuo Akiyama: 'Study concerning the colours of tactile blocks for the visually handicapped-visibility for the visually handicapped and scenic congruence for those with ordinary sight and vision'. *JIPEA World Congress*. 2000: pp. 453–462 (cit. on pp. 2, 22).

[Sal07] Salvador, Stan and Philip Chan: 'Toward Accurate Dynamic Time Warping in Linear Time and Space'. *Intell. Data Anal.* (2007), vol. 11(5): pp. 561–580 (cit. on p. 38).

[Sat15] Sato, Munehiko, Shigeo Yoshida, Alex Olwal, Boxin Shi, Atsushi Hiyama, Tomohiro Tanikawa, Michitaka Hirose, and Ramesh Raskar: 'SpecTrans: Versatile Material Classification for Interaction with Textureless, Specular and Transparent Surfaces'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* CHI '15. Seoul, Republic of Korea: ACM, 2015: pp. 2191–2200 (cit. on p. 117).

[Sav10] Savage, Neil: 'Cycling through Data'. *Commun. ACM* (2010), vol. 53(9): pp. 16–17 (cit. on pp. 92, 94).

[Sav17] Saviot, Léa, Frederik Brudy, and Steven Houben: 'WRISTBAND.IO: Expanding Input and Output Spaces of a Smartwatch'. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* CHI EA '17. Denver, Colorado, USA: Association for Computing Machinery, 2017: pp. 2025–2033 (cit. on p. 72).

[Saw00] Sawhney, Nitin and Chris Schmandt: 'Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments'. *ACM transactions on Computer-Human interaction (TOCHI)* (2000), vol. 7(3): pp. 353–383 (cit. on p. 155).

[Say86] Sayers, M.: 'The International Road Roughness Experiment (IRRE) : establishing correlation and a calibration standard for measurements'. 1986 (cit. on p. 93).

[Sch93] Schafer, R Murray: *The soundscape: Our sonic environment and the tuning of the world.* Simon and Schuster, 1993 (cit. on p. 155).

[Sch02] Schmidt, Albrecht, Martin Strohbach, Kristof van Laerhoven, et al.: 'Ubiquitous interaction—using surfaces in everyday environments as pointing devices'. *ERCIM Workshop on User Interfaces for All.* Springer. 2002: pp. 263–279 (cit. on p. 19).

[Sch11] Schmidt, Robert F, Florian Lang, and Manfred Heckmann: *Physiologie des menschen: mit pathophysiologie.* Springer-Verlag, 2011 (cit. on p. 10).

[Sch16] Schneegass, Stefan and Alexandra Voit: 'GestureSleeve: using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches'. Sept. 2016: pp. 108–115 (cit. on p. 72).

[Sch19a] Schneider, Kurt and Busch, Melanie and Karras, Oliver and Schrapel, Maximilian and Rohs, Michael: 'Refining vision videos'. *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer. 2019: pp. 135–150 (cit. on p. 267).

[Sch08a] Schöning, Johannes, Peter Brandl, Florian Daiber, Florian Echtler, Otmar Hilliges, Jonathan Hook, Markus Löchtefeld, Nima Motamedi, Laurence Muller, Patrick Olivier, et al.: 'Multi-touch surfaces: A technical guide'. *Johannes Schöning, Institute for Geoinformatics University of Münster, Technical Report TUM-I0833* (2008), vol. (cit. on pp. 18, 20).

[Sch10] Schöning, Johannes, Jonathan Hook, Tom Bartindale, Dominik Schmidt, Patrick Oliver, Florian Echtler, Nima Motamedi, Peter Brandl, and Ulrich von Zadow: 'Building Interactive Multi-touch Surfaces'. *Tabletops - Horizontal Interactive Displays*. Ed. by Müller-Tomfelde, Christian. London: Springer London, 2010: pp. 27–49 (cit. on pp. 18–20).

[Sch22a] Schrapel, Maximilian, Dennis Grannemann, and Michael Rohs: 'Sign H3re: Symbol and X-Mark Writer Identification Using Audio and Motion Data from a Digital Pen'. *Proceedings of Mensch Und Computer 2022*. MuC '22. Darmstadt, Germany: Association for Computing Machinery, 2022: pp. 209–218 (cit. on pp. j, 53, 267, 269).

[Sch21] Schrapel, Maximilian & Etgeton, Philipp & Rohs, Michael: 'SpectroPhone: Enabling Material Surface Sensing with Rear Camera and Flashlight LEDs'. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021 (cit. on pp. j, 115, 268).

[Sch20a] Schrapel, Maximilian & Herzog, Florian & Ryll, Steffen & Rohs, Michael: 'Watch My Painting: The Back of the Hand as a Drawing Space for Smartwatches'. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, 2020: pp. 1–10 (cit. on pp. i, 56, 69, 267).

[Sch20b] Schrapel, Maximilian & Liebers, Jonathan & Rohs, Michael & Schneegass, Stefan: 'Skiables: Towards a Wearable System Mounted on a Ski Boot for Measuring Slope Conditions'. *19th International Conference on Mobile and Ubiquitous Multimedia*. MUM 2020. Essen, Germany: Association for Computing Machinery, 2020: pp. 320–322 (cit. on pp. i, 91, 268).

[Sch20c] Schrapel, Maximilian & Schulz, Thilo & Rohs, Michael: 'Augmenting Public Bookcases to Support Book Sharing'. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '20. Oldenburg, Germany: Association for Computing Machinery, 2020 (cit. on pp. j, 129, 267, 269).

[Sch18] Schrapel, Maximilian & Stadler, Max-Ludwig & Rohs, Michael: 'Pentelligence: Combining Pen Tip Motion and Writing Sounds for Handwritten Digit Recognition'. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2018: pp. 1–11 (cit. on pp. i, 31, 55–58, 60, 64, 65, 72, 267).

[Sch22b] Schrapel, Maximilian and Happe, Janko and Rohs, Michael: 'EnvironZen: Immersive Soundscapes via Augmented Footstep Sounds in Urban Areas'. *i-com* (2022), vol. 21(2): pp. 219–237 (cit. on pp. j, 153, 267).

[Sch08b] Schreiner, Keri: 'Uniting the paper and digital worlds'. *IEEE computer graphics and applications* (2008), vol. 28(6): pp. 6–10 (cit. on p. 21).

[Sch19b] Schulz, Thilo: 'Entwicklung einer mobilen AR-Anwendungsunterstützung der visuellen Suche'. Master Thesis. Leibniz Universität Hannover, 2019 (cit. on pp. k, 129).

[Seb19] Sebkhi, N., N. Sahadat, S. Hersek, A. Bhavsar, S. Siahpoushan, M. Ghoovanloo, and O. T. Inan: 'A Deep Neural Network-Based Permanent Magnet Localization for Tongue Tracking'. *IEEE Sensors Journal* (2019), vol. 19(20): pp. 9324–9331 (cit. on p. 72).

[Sei17] Seiça, Mariana, Buga Lopes, Pedro Martins, and Amílcar Cardoso: 'Sonifying Twitter's Emotions Through Music'. Sept. 2017 (cit. on pp. 155, 156).

[Sei] Seiffert, Udo: *HawkSpex Mobile*. Accessed: 2021-02-02 (cit. on pp. 117, 120).

[Sen09] Seniuk, Andrew and Dorothea Blostein: 'Pen Acoustic Emissions for Text and Gesture Recognition'. *Proceedings of the 2009 10th International Conference on Document Analysis and Recognition*. ICDAR '09. Washington, DC, USA: IEEE Computer Society, 2009: pp. 872–876 (cit. on p. 34).

[Ser11] Serafin, S., L. Turchet, and R. Nordahl: 'Auditory feedback in a multimodal balancing task: Walking on a virtual plank'. 2011 (cit. on pp. 13, 156).

[Sha20] Shadan Sadeghian & Maximilian Schrapel et al.: *Workshop on Designing Technologies for Future Forms of Sustainable Mobility. MobileHCI'20, Oldenburg, Germany.* 2020 (cit. on pp. i, 91, 92, 268).

[Sha49] Shannon, Claude Elwood: 'Communication in the presence of noise'. *Proceedings of the IRE* (1949), vol. 37(1): pp. 10–21 (cit. on p. 60).

[She22] Shen, Vivian, Craig Shultz, and Chris Harrison: 'Mouth Haptics in VR Using a Headset Ultrasound Phased Array'. *CHI Conference on Human Factors in Computing Systems.* CHI '22. New Orleans, LA, USA: Association for Computing Machinery, 2022 (cit. on p. 27).

[She06] Shepard, Mark: 'Tactical sound garden [tsg] toolkit'. *3rd international workshop on mobile music technology.* 2006 (cit. on pp. 155–157).

[She07] Shepard, Mark: 'Tactical Sound Garden Toolkit'. *ACM SIGGRAPH 2007 Art Gallery.* SIGGRAPH '07. San Diego, California: Association for Computing Machinery, 2007: p. 219 (cit. on p. 155).

[Shi20] Shi, Yilei, Haimo Zhang, Kaixing Zhao, Jiashuo Cao, Mengmeng Sun, and Suranga Nanayakkara: 'Ready, steady, touch! sensing physical contact with a finger-mounted IMU'. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2020), vol. 4(2): pp. 1–25 (cit. on p. 26).

[Shi16] Shibata, Tomoki, Daniel Afergan, Danielle Kong, Beste F. Yuksel, Scott MacKenzie, and Robert J.K. Jacob: 'Text Entry for Ultra-Small Touchscreens Using a Fixed Cursor and Movable Keyboard'. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* CHI EA '16. San Jose, California, USA: ACM, 2016: pp. 3770–3773 (cit. on p. 72).

[Shi08] Shinn-Cunningham, Barbara G. and Virginia Best: 'Selective Attention in Normal and Impaired Hearing'. *Trends in Amplification* (2008), vol. 12(4). PMID: 18974202: pp. 283–299 (cit. on p. 154).

[Shn81] Shneiderman, Ben: 'Direct manipulation: A step beyond programming languages'. *Proceedings of the Joint Conference on Easier and More Productive Use of Computer Systems.(Part-II): Human Interface and the User Interface-Volume 1981.* 1981: p. 143 (cit. on p. 17).

[Shn82] Shneiderman, Ben: 'The future of interactive systems and the emergence of direct manipulation'. *Behaviour & Information Technology* (1982), vol. 1(3): pp. 237–256 (cit. on p. 17).

[Sik18] Sikora, Marjan, Mladen Russo, Jurica undefinederek, and Ante Jurčević: 'Soundscape of an Archaeological Site Recreated with Audio Augmented Reality'. *ACM Trans. Multimedia Comput. Commun. Appl.* (July 2018), vol. 14(3) (cit. on pp. 155, 156).

[Smi02] Smith, Allan M, Geneviéve Gosselin, and Bryan Houde: 'Deployment of fingertip forces in tactile exploration'. *Experimental brain research* (2002), vol. 147(2): pp. 209–218 (cit. on p. 15).

[Smi18] Smith, Daniel A: 'Vote-by-mail ballots cast in florida'. *ACLU of Florida. URL: https://www. aclufl. org/en/press-releases/aclu-report-finds-varying-rates-rejectedvote-mail-ballots-across-florida-counties* (2018), vol. (cit. on p. 54).

[Sno15] Snow, Marianne: 'Little Free Libraries: A Call for Research into the Tiny Book Depositories'. *Children and Libraries* (Dec. 2015), vol. 13: p. 30 (cit. on p. 133).

[Soe21] Soelistio, Elizabeth Ann, Rafael Edwin Hananto Kusumo, Zevira Varies Martan, and Edy Irwansyah: 'A Review of Signature Recognition Using Machine Learning'. *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*. Vol. 1. IEEE. 2021: pp. 219–223 (cit. on p. 55).

[Sol18] Soliman, Mohamed, Franziska Mueller, Lena Hegemann, Joan Sol Roo, Christian Theobalt, and Jürgen Steimle: 'Fingerinput: Capturing expressive single-hand thumb-to-finger microgestures'. *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*. 2018: pp. 177–187 (cit. on p. 25).

[Son18] Son, Hyungki, Hyunjae Gil, Sangkyu Byeon, Sang-Youn Kim, and Jin Ryong Kim: 'Realwalk: Feeling ground surfaces while walking in virtual reality'. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018: pp. 1–4 (cit. on pp. 27, 156).

[Sou69] Southworth, Michael: 'The Sonic Environment of Cities'. *Environment and Behavior* (1969), vol. 1(1): pp. 49–70 (cit. on pp. 9, 154).

[Spe12] Spelmezan, Daniel: 'An Investigation into the Use of Tactile Instructions in Snowboarding'. *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '12. San Francisco, California, USA: Association for Computing Machinery, 2012: pp. 417–426 (cit. on p. 95).

[Spe09] Spelmezan, Daniel, Mareike Jacobs, Anke Hilgers, and Jan Borchers: 'Tactile Motion Instructions for Physical Activities'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '09. Boston, MA, USA: Association for Computing Machinery, 2009: pp. 2243–2252 (cit. on p. 95).

[Spe10] Sperber, Matthias, Martin Klinkigt, Koichi Kise, Masakazu Iwamura, Benjamin Adrian, and Andreas Dengel: 'Handwriting reconstruction for a camera pen using random dot patterns'. *2010 12th International Conference on Frontiers in Handwriting Recognition*. IEEE. 2010: pp. 160–165 (cit. on p. 56).

[Sre17] Srednicki, Nicole Rachel: *Exploring a Change in Meditation Practice Using Noise-Canceling Headphones*. Georgetown University, 2017 (cit. on p. 154).

[Sri17] Sridhar, Srinath, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring: 'WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor'. May 2017: pp. 3891–3902 (cit. on pp. 25, 72).

[Sri08] Srihari, Sargur, Chen Huang, and Harish Srinivasan: 'On the discriminability of the handwriting of twins'. *Journal of Forensic Sciences* (2008), vol. 53(2): pp. 430–446 (cit. on pp. 54, 66).

[Sri14] Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov: 'Dropout: A Simple Way to Prevent Neural Networks from Overfitting'. *Journal of Machine Learning Research* (2014), vol. 15: pp. 1929–1958 (cit. on p. 41).

[Gmb22] GmbH, STABILO International: *DigiVision*. https://stabilodigital.com/. 2022 (cit. on pp. 21, 56, 186).

[Sta17] Stadler, Max-Ludwig: 'Deep Learning zur Handschriftenerkennung mittels Audio- und Bewegungsdaten von Stiften'. Master Thesis. Leibniz Universität Hannover, 2017 (cit. on pp. k, 31).

[Sta21] Stauffer, Michael, Paul Maergner, Andreas Fischer, and Kaspar Riesen: 'A survey of state of the art methods employed in the offline signature verification process'. *New Trends in Business Information Systems and Technology*. Springer, 2021: pp. 17–30 (cit. on p. 55).

[Ste07] Steimle, Jürgen, Iryna Gurevych, and Max Mühlhäuser: 'Notetaking in University Courses and its Implications for eLearning Systems'. *DeLFI 2007: 5. e-Learning Fachtagung Informatik*. Ed. by Eibl, Christian, Johannes Magenheim, Sigrid Schubert, and Martin

Wessner. Bonn: Gesellschaft für Informatik, 2007: pp. 45–56 (cit. on p. 32).

[Sti03] Stinson, Monique and Chandra Bhat: 'Commuter Bicyclist Route Choice: Analysis Using a Stated Preference Survey'. *Transportation Research Record* (Jan. 2003), vol. 1828: pp. 107–115 (cit. on p. 93).

[Str05] Strauss, Olaf: 'The retinal pigment epithelium in visual function'. *Physiological reviews* (2005), vol. 85(3): pp. 845–881 (cit. on p. 10).

[Str15] Strohmeier, Paul, Jesse Burstyn, and Roel Vertegaal: 'Effects of Display Sizes on a Scrolling Task Using a Cylindrical Smartwatch'. *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '15. Copenhagen, Denmark: ACM, 2015: pp. 846–853 (cit. on p. 72).

[Suh16] Suh, Jihoon: 'Veri-Pen: A Pen-Based Identification Through Natural Biometrics Extraction'. CHI EA '16. San Jose, California, USA: Association for Computing Machinery, 2016: pp. 134–139 (cit. on pp. 56, 66, 67).

[Sun16] Sun, Ziwen, Yao Wang, Gang Qu, and Zhiping Zhou: 'A 3-D Hand Gesture Signature Based Biometric Authentication System for Smartphones'. *Security and Communication Networks* (July 2016), vol. 9 (cit. on p. 55).

[Sun19] Sundaram, Subramanian, Petr Kellnhofer, Yunzhu Li, Jun-Yan Zhu, Antonio Torralba, and Wojciech Matusik: 'Learning the signatures of the human grasp using a scalable tactile glove'. *Nature* (2019), vol. 569(7758): pp. 698–702 (cit. on p. 27).

[Sut68] Sutherland, Ivan E: 'A head-mounted three dimensional display'. *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968: pp. 757–764 (cit. on p. 23).

[Sut64] Sutherland, Ivan E: 'Sketchpad a man-machine graphical communication system'. *Simulation* (1964), vol. 2(5): R–3 (cit. on p. 17).

[Taj15] Tajadura-Jiménez, Ana, Maria Basia, Ophelia Deroy, Merle Fairhurst, Nicolai Marquardt, and Nadia Bianchi-Berthouze: 'As Light as Your Footsteps: Altering Walking Sounds to Change Perceived Body Weight, Emotional State and Gait'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: Association for Computing Machinery, 2015: pp. 2943–2952 (cit. on pp. 13, 27, 156).

[Tak10] TAKEUCHI, YUICHIRO: 'Gilded gait: reshaping the urban experience with augmented footsteps'. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. 2010: pp. 185–188 (cit. on pp. 27, 156).

[Tal14] TALKER, LIOR and YAEL MOSES: 'Viewpoint-independent book spine segmentation'. *IEEE Winter Conference on Applications of Computer Vision* (2014), vol.: pp. 453–460 (cit. on p. 134).

[Tay10] TAYLOR, M and CA FAIRFIELD: 'IntelliBike: condition monitoring of our cycling infrastructure'. *Bicycle and Motorcycle Dynamics Symposium on the Dynamics and Control of Single Track Vehicles, 20th–22nd October, Delft, The Netherlands*. 2010 (cit. on p. 94).

[Tec22] TECHNOLOGIES, UNITY: *Unity 3D 2020 LTS*. Online; accessed 1 March 2022. 2022 (cit. on p. 176).

[Tel02] TELECOMMUNICATIONS, NATIONAL and UNITED STATES DEPARTMENT OF COMMERCE INFORMATION ADMINISTRATION (NITA): *Product Recall Exception to the Electronic Signatures in Global and National Commerce Act*. https://www.ntia.doc.gov/files/ntia/publications/fr_esign_recall_020924.pdf. Accessed: 2022-01-03. 2002 (cit. on p. 54).

[Ten00] TENNENHOUSE, DAVID: 'Proactive Computing'. *Commun. ACM* (May 2000), vol. 43(5): pp. 43–50 (cit. on p. 16).

[Tey17] TEYSSIER, MARC, GILLES BAILLY, and ÉRIC LECOLINET: 'VersaPen: Exploring Multimodal Interactions with a Programmable Modular Pen'. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '17. Denver, Colorado, USA: Association for Computing Machinery, 2017: pp. 377–380 (cit. on p. 56).

[Til07] TILAHUN, NEBIYOU Y, DAVID M LEVINSON, and KEVIN J KRIZEK: 'Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey'. *Transportation Research Part A: Policy and Practice* (2007), vol. 41(4): pp. 287–301 (cit. on p. 93).

[Tol21] TOLOSANA, RUBEN, RUBEN VERA-RODRIGUEZ, JULIAN FIERREZ, and JAVIER ORTEGA-GARCIA: 'DeepSign: Deep on-line signature verification'. *IEEE Transactions on Biometrics, Behavior, and Identity Science* (2021), vol. 3(2): pp. 229–239 (cit. on p. 55).

[Tre80] TREISMAN, ANNE M. and GARRY GELADE: 'A feature-integration theory of attention'. *Cognitive Psychology* (1980), vol. 12(1): pp. 97–136 (cit. on p. 133).

[Tri21] TRIGGS, ROBERT: *Android's Bluetooth latency needs a serious overhaul*. Online; accessed 12 June 2022. 2021 (cit. on p. 162).

[Tsa16a] TSAI, HSIN-RUEY, MIN-CHIEH HSIU, JUI-CHUN HSIAO, LEE-TING HUANG, MIKE CHEN, and YI-PING HUNG: 'TouchRing: Subtle and Always-Available Input Using a Multi-Touch Ring'. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '16. Florence, Italy: Association for Computing Machinery, 2016: pp. 891–898 (cit. on p. 72).

[Tsa16b] TSAI, HSIN-RUEY, CHENG-YUAN WU, LEE-TING HUANG, and YI-PING HUNG: 'ThumbRing: Private Interactions Using One-Handed Thumb Motion Input on Finger Segments'. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. MobileHCI '16. Florence, Italy: Association for Computing Machinery, 2016: pp. 791–798 (cit. on p. 72).

[Tsa10] TSAI, SAM S., DAVID CHEN, VIJAY CHANDRASEKHAR, GABRIEL TAKACS, NGAI-MAN CHEUNG, RAMAKRISHNA VEDANTHAM, RADEK GRZESZCZUK, and BERND GIROD: 'Mobile Product Recognition'. *Proceedings of the 18th ACM International Conference on Multimedia*. MM '10. Firenze, Italy: ACM, 2010: pp. 1587–1590 (cit. on p. 134).

[Tur15] TURCHET, LUCA, IVAN CAMPONOGARA, and PAOLA CESARI: 'Interactive footstep sounds modulate the perceptual-motor aftereffect of treadmill walking'. *Experimental brain research* (2015), vol. 233(1): pp. 205–214 (cit. on pp. 13, 156).

[Tur13a] TURCHET, LUCA and STEFANIA SERAFIN: 'Investigating the amplitude of interactive footstep sounds and soundscape reproduction'. *Applied Acoustics* (2013), vol. 74(4): pp. 566–574 (cit. on pp. 156, 172, 173).

[Tur13b] TURCHET, LUCA, STEFANIA SERAFIN, and PAOLA CESARI: 'Walking Pace Affected by Interactive Sounds Simulating Stepping on Different Terrains'. *ACM Trans. Appl. Percept.* (2013), vol. 10(4) (cit. on pp. 2, 13, 156, 157, 172).

[Twi] TWITTER: *Twitter API v2*. `https://developer.twitter.com/en/docs/api-reference-index`. Accessed: 2022-02-23 (cit. on p. 125).

[Uch09] UCHIDA, SEIICHI, KATSUHIRO ITOU, MASAKAZU IWAMURA, SHINICHIRO OMACHI, and KOICHI KISE: 'On a possibility of pen-tip camera for the reconstruction of handwritings'. *Proc. CBDAR2009* (2009), vol.: pp. 119–126 (cit. on p. 21).

[van16] VAN HERWIJNEN, A., M. HECK, and J. SCHWEIZER: 'Forecasting snow avalanches using avalanche activity data obtained through seismic monitoring'. *Cold Regions Science and Technology* (2016), vol. 132: pp. 68–80 (cit. on p. 95).

[Van20] VAN RENTERGHEM, TIMOTHY, KRIS VANHECKE, KARLO FILIPAN, KANG SUN, TOON DE PESSEMIER, BERT DE COENSEL, WOUT JOSEPH, and DICK BOTTELDOOREN: 'Interactive soundscape augmentation by natural sounds in a noise polluted urban park'. *Landscape and Urban Planning* (2020), vol. 194: p. 103705 (cit. on pp. 155, 156).

[Vaz12] VAZQUEZ-ALVAREZ, YOLANDA, IAN OAKLEY, and STEPHEN A. BREWSTER: 'Auditory Display Design for Exploration in Mobile Audio-Augmented Reality'. *Personal Ubiquitous Comput.* (Dec. 2012), vol. 16(8): pp. 987–999 (cit. on pp. 155–157).

[Vir20] VIRTANEN, PAULI et al.: 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python'. *Nature Methods* (2020), vol. 17: pp. 261–272 (cit. on p. 105).

[Vuj16] VUJIČIC, TIJANA, TRIPO MATIJEVIC, JELENA LJUCOVIC, ADIS BALOTA, and ZORAN ŠEVARAC: 'Comparative Analysis of Methods for Determining Number of Hidden Neurons in Artificial Neural Network'. *Proceeding of the Central European Conference on Information and Intelligent Systems* (2016), vol. 27: pp. 219–223 (cit. on p. 41).

[Com22] COMPANY, WACOM: *Wacom Inkling*. http://inkling.wacom.eu. 2022 (cit. on pp. 21, 32, 56).

[Wag20] WAGE, O., U. FEUERHAKE, C. KOETSIER, A. PONICK, N. SCHILD, T. BEENING, and S. DARE: *Ride Vibrations: Towards Comfort-Based Bicycle Navigation*. English. 2020 (cit. on pp. 92–94, 97, 102, 110).

[Wal12] WALKER, GEOFF: 'A review of technologies for sensing contact location on the surface of a display'. *Journal of the Society for Information Display* (2012), vol. 20(8): pp. 413–440 (cit. on p. 18).

[Wan18] WANG, EDWARD JAY, JUNYI ZHU, MOHIT JAIN, TIEN-JUI LEE, ELLIOT SABA, LAMA NACHMAN, and SHWETAK N. PATEL: 'Seismo: Blood Pressure Monitoring Using Built-in Smartphone Accelerometer and Camera'. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018: pp. 1–9 (cit. on p. 116).

[Wan17] Wang, Edward Jay, Junyi Zhu, William Li, Rajneil Rana, and Shwetak Patel: 'HemaApp IR: Noninvasive Hemoglobin Measurement Using Unmodified Smartphone Cameras and Built-in LEDs'. UbiComp '17. Maui, Hawaii: Association for Computing Machinery, 2017: pp. 305–308 (cit. on p. 116).

[Wan12] Wang, Jeen-Shing and Fang-Chen Chuang: 'An Accelerometer-Based Digital Pen With a Trajectory Recognition Algorithm for Handwritten Digit and Gesture Recognition'. *IEEE Trans. Industrial Electronics* (2012), vol. 59(7): pp. 2998–3007 (cit. on pp. 33, 34, 39, 56).

[Wan16a] Wang, Qinglong, Xiangshi Ren, Sayan Sarcar, and Xiaoying Sun: 'EV-Pen: Leveraging Electrovibration Haptic Feedback in Pen Interaction'. *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ISS '16. Niagara Falls, Ontario, Canada: Association for Computing Machinery, 2016: pp. 57–66 (cit. on pp. 22, 56).

[Wan16b] Wang, Yi, Xiaohu Liu, Peng Chen, Nhung Thi Tran, Jinling Zhang, Wei Sheng Chia, Souhir Boujday, and Bo Liedberg: 'Smartphone spectrometer for colorimetric biosensing'. (Apr. 2016), vol. (cit. on p. 117).

[Wan21] Wang, Zhengjie, Naisheng Zhou, Fang Chen, Xiaoxue Feng, Fei Liu, Yinjing Guo, and Da Chen: 'Smart_Auth: User Identity Authentication Based on Smartphone Motion Sensors'. *2021 6th International Conference on Image, Vision and Computing (ICIVC)*. IEEE. 2021: pp. 480–485 (cit. on p. 55).

[Wei21] Wei, Zhixiang, Song Yang, Yadong Xie, Fan Li, and Bo Zhao: 'SVSV: Online handwritten signature verification based on sound and vibration'. *Information Sciences* (2021), vol. 572: pp. 109–125 (cit. on pp. 55, 56).

[Wei15] Weigel, Martin, Tong Lu, Gilles Bailly, Antti Oulasvirta, Carmel Majidi, and Jürgen Steimle: 'Iskin: flexible, stretchable and visually customizable on-body touch sensors for mobile computing'. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015: pp. 2991–3000 (cit. on p. 27).

[Wei14] Weigel, Martin, Vikram Mehta, and Jürgen Steimle: 'More than touch: understanding how people use skin as an input surface for mobile computing'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014: pp. 179–188 (cit. on p. 25).

[Wei91] Weiser, Mark: ‘The computer for the 21st century’. *Scientific American* (1991), vol. 265(3): pp. 94–104 (cit. on pp. 1, 16, 17, 92, 181).

[Wei11] Weiss, Malte, Chat Wacharamanotham, Simon Voelker, and Jan Borchers: ‘FingerFlux: Near-surface Haptic Feedback on Tabletops’. *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. UIST ’11. Santa Barbara, California, USA: ACM, 2011: pp. 615–620 (cit. on p. 72).

[Wel91] Wellner, Pierre: ‘The DigitalDesk calculator: tangible manipulation on a desk top display’. *Proceedings of the 4th annual ACM symposium on User interface software and technology*. 1991: pp. 27–33 (cit. on pp. 18, 23, 28).

[Wen10] Wendt, Gunnar, Franz Faul, Vebjørn Ekroll, and Rainer Mausfeld: ‘Disparity, motion, and color information improve gloss constancy performance’. *Journal of vision* (2010), vol. 10(9): pp. 7–7 (cit. on p. 11).

[Whi15] Whittaker, Edmund Taylor: ‘XVIII.—On the functions which are represented by the expansions of the interpolation-theory’. *Proceedings of the Royal Society of Edinburgh* (1915), vol. 35: pp. 181–194 (cit. on p. 60).

[Wig07] Wigdor, Daniel, Clifton Forlines, Patrick Baudisch, John Barnwell, and Chia Shen: ‘Lucid Touch: A See-through Mobile Device’. *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*. UIST ’07. Newport, Rhode Island, USA: Association for Computing Machinery, 2007: pp. 269–278 (cit. on pp. 20, 32, 70).

[Wij18] Wijerathne, Nipun, Sanjana Kadaba Viswanath, Marakkalage Sumudu Hasala, Victoria Beltran, Chau Yuen, and Hock Beng Lim: ‘Towards comfortable cycling: A practical approach to monitor the conditions in cycling paths’. *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. 2018: pp. 778–783 (cit. on pp. 93, 94).

[Wik09] Wikimedia Commons Lars Chittka, Axel Brockmann: *A diagram of the anatomy of the human ear*. File: Anatomy$_of_the_Human_Ear.svg$. 2009. url: https://de.wikipedia.org/wiki/File:Anatomy_of_the_Human_Ear.svg (cit. on p. 13).

[Wik17] Wikimedia Commons Dr. Hegasy, Guido4: *06 Hegasy Skin Layers Receptors Wiki EN CCBYSA*. File: 06$_Hegasy_Skin_Layers_Receptors_Wiki_EN_CCBYSA.png$. 2017. url: https://commons.wikimedia.org/wiki/File:06_Hegasy_Skin_Layers_Receptors_Wiki_EN_CCBYSA.png (cit. on p. 16).

[Wik13] WIKIMEDIA COMMONS, HOLLY FISCHER: *Three Main Layers of the Eye*. File: Three$_M ain_L ayers_o f_t he_E ye.png$. 2013. URL: https://commons.wikimedia.org/wiki/File:Three_Main_Layers_of_the_Eye.png (cit. on p. 11).

[Wik04] WIKIMEDIA COMMONS, OARIH ROPSHKOW: *Cochlea-crosssection*. File: Cochlea-crosssection.png. 2004. URL: https://commons.wikimedia.org/wiki/File:Cochlea-crosssection.png (cit. on p. 13).

[Wil20] WILLICH, JULIUS von, MARTIN SCHMITZ, FLORIAN MÜLLER, DANIEL SCHMITT, and MAX MÜHLHÄUSER: 'Podoportation: Foot-Based Locomotion in Virtual Reality'. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2020: pp. 1–14 (cit. on pp. 13, 156).

[Wil10] WILSON, ANDREW D: 'Using a depth camera as a touch sensor'. *ACM international conference on interactive tabletops and surfaces*. 2010: pp. 69–72 (cit. on pp. 20, 26).

[Wil21] WILSON, RW: 'Active Surface Technology For The Lunar Crater Radio Telescope On The Far-side Of The Moon'. *American Astronomical Society Meeting Abstracts*. Vol. 53. 6. 2021: pp. 309–02 (cit. on p. 17).

[Win10] WINTERS, MEGHAN, KAY TESCHKE, MICHAEL GRANT, ELEANOR M. SETTON, and MICHAEL BRAUER: 'How Far Out of the Way Will We Travel?: Built Environment Influences on Route Selection for Bicycle and Car Travel'. *Transportation Research Record* (2010), vol. 2190(1): pp. 1–10 (cit. on pp. 92, 93).

[Wir95] WIRTZ, B.: 'Stroke-based time warping for signature verification'. *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 1995: 179–182 vol.1 (cit. on p. 55).

[Wit18] WITHANA, ANUSHA, DANIEL GROEGER, and JÜRGEN STEIMLE: 'Tacttoo: A Thin and Feel-Through Tattoo for On-Skin Tactile Output'. *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. UIST '18. Berlin, Germany: Association for Computing Machinery, 2018: pp. 365–378 (cit. on p. 27).

[Wit98] WITMER, BOB G. and MICHAEL J. SINGER: 'Measuring Presence in Virtual Environments: A Presence Questionnaire'. *Presence: Teleoperators and Virtual Environments* (June 1998), vol. 7(3): pp. 225–240 (cit. on p. 174).

[Won17] WONG, PUI, ZHU KENING, and HONGBO FU: 'fingerT9 : leveraging thumb-to-finger interaction for one-handed text entry on smartwatches'. Nov. 2017: pp. 1–3 (cit. on p. 72).

[Woo16] Woods, Terri, Sarah Reed, Sherry Hsi, John Woods, and Michael Woods: 'Pilot Study Using the Augmented Reality Sandbox to Teach Topographic Maps and Surficial Processes in Introductory Geology Labs'. *Journal of Geoscience Education* (Aug. 2016), vol. 64: pp. 199–214 (cit. on p. 18).

[Woz17] Wozniak, Paweł W., Anton Fedosov, Eleonora Mencarini, and Kristina Knaving: 'Soil, Rock, and Snow: On Designing for Information Sharing in Outdoor Sports'. *Proceedings of the 2017 Conference on Designing Interactive Systems*. DIS '17. Edinburgh, United Kingdom: Association for Computing Machinery, 2017: pp. 611–623 (cit. on pp. 94, 95).

[Wu17] Wu, Po-Chen, Robert Wang, Kenrick Kin, Christopher Twigg, Shangchen Han, Ming-Hsuan Yang, and Shao-Yi Chien: 'DodecaPen: Accurate 6DoF Tracking of a Passive Stylus'. *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. UIST '17. Québec City, QC, Canada: Association for Computing Machinery, 2017: pp. 365–374 (cit. on p. 72).

[Wu20] Wu, Z. and H. Lin: 'IntelliBike: An Intelligent Bicycle with Automatic Shifting Based on Sensing and Riding Analysis'. *2020 IEEE/SICE International Symposium on System Integration (SII)*. 2020: pp. 731–736 (cit. on p. 94).

[Wüs18] Wüst, Karl and Arthur Gervais: 'Do you need a blockchain?' *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. IEEE. 2018: pp. 45–54 (cit. on p. 65).

[Xia15] Xia, Haijun, Tovi Grossman, and George Fitzmaurice: 'NanoStylus: Enhancing Input on Ultra-Small Displays with a Finger-Mounted Stylus'. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. UIST '15. Charlotte, NC, USA: ACM, 2015: pp. 447–456 (cit. on pp. 20, 72).

[Ya-99] Ya-Xian, Zhen, Takaki Suetake, and Hachiro Tagami: 'Number of cell layers of the stratum corneum in normal skin–relationship to the anatomical location on the body, age, sex and physical parameters'. *Archives of dermatological research* (1999), vol. 291(10): pp. 555–559 (cit. on p. 14).

[Xia16] Xiao, Robert, Scott Hudson, and Chris Harrison: 'Direct: Making touch tracking on ordinary surfaces practical with hybrid depth-infrared sensing'. *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. 2016: pp. 85–94 (cit. on p. 26).

[Xia14] Xiao, Robert, Gierad Laput, and Chris Harrison: 'Expanding the Input Expressivity of Smartwatches with Mechanical Pan, Twist, Tilt and Click'. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: ACM, 2014: pp. 193–196 (cit. on p. 72).

[Xia18] Xiao, Robert, Julia Schwarz, Nick Throm, Andrew D Wilson, and Hrvoje Benko: 'MRTouch: Adding touch input to head-mounted mixed reality'. *IEEE transactions on visualization and computer graphics* (2018), vol. 24(4): pp. 1653–1660 (cit. on p. 26).

[Xio21] Xiong, Jianghao, En-Lin Hsiang, Ziqian He, Tao Zhan, and Shin-Tson Wu: 'Augmented reality and virtual reality displays: emerging technologies and future perspectives'. *Light: Science & Applications* (2021), vol. 10(1): pp. 1–30 (cit. on pp. 24, 185).

[Xu11] Xu, Cheng, Ali Israr, Ivan Poupyrev, Olivier Bau, and Chris Harrison: 'Tactile display for the visually impaired using TeslaTouch'. *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 2011: pp. 317–322 (cit. on p. 23).

[Xu20] Xu, Songlin, Zhiyuan Wu, Shunhong Wang, Rui Fan, and Nan Lin: 'Hydrauio: Extending Interaction Space on the Pen through Hydraulic Sensing and Haptic Output'. *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology*. UIST '20 Adjunct. Virtual Event, USA: Association for Computing Machinery, 2020: pp. 43–45 (cit. on p. 56).

[Yan17] Yang, X., D. He, W. Huang, A. Ororbia, Z. Zhou, D. Kifer, and C. L. Giles: 'Smart Library: Identifying Books on Library Shelves Using Supervised Deep Learning for Scene Text Reading'. *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. 2017: pp. 1–4 (cit. on p. 134).

[Yan09] Yanikoglu, Berrin and Alisher Kholmatov: 'Online Signature Verification Using Fourier Descriptors'. *Journal on Advances in Signal Processing* (Dec. 2009), vol. (cit. on pp. 59, 64).

[Yeo17] Yeo, Hui-Shyong, Juyoung Lee, Andrea Bianchi, David Harris-Birtill, and Aaron Quigley: 'SpeCam: sensing surface color and material with the front-facing camera of a mobile device'. Sept. 2017: pp. 1–9 (cit. on pp. 4, 115–118, 120–122, 127).

[Yi17] Yi, Xin, Chun Yu, Weijie Xu, Xiaojun Bi, and Yuanchun Shi: 'COMPASS: Rotational Keyboard on Non-Touch Smartwatches'. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: ACM, 2017: pp. 705–715 (cit. on p. 72).

[Yil15] Yildirim, Caglar and Ana-Paula Correia: 'Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire'. *Computers in Human Behavior* (2015), vol. 49: pp. 130–137 (cit. on p. 126).

[Yin07] Ying, Hong, C. Silex, A. Schnitzer, S. Leonhardt, and M. Schiek: 'Automatic Step Detection in the Accelerometer Signal'. *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*. Ed. by Leonhardt, Steffen, Thomas Falck, and Petri Mähönen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: pp. 80–85 (cit. on p. 159).

[Yon10] Yoneyama, Takeshi, Motoki Kitade, and Kazutaka Osada: 'Investigation on the ski-snow interaction in a carved turn based on the actual measurement'. *Procedia Engineering* (June 2010), vol. 2: pp. 2901–2906 (cit. on pp. 95–97).

[Yoo16a] Yoon, Sang Ho, Ke Huo, and Karthik Ramani: 'TMotion: Embedded 3D Mobile Input Using Magnetic Sensing Technique'. *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. TEI '16. Eindhoven, Netherlands: ACM, 2016: pp. 21–29 (cit. on pp. 72, 73, 75, 77, 78).

[Yoo16b] Yoon, Sang Ho, Yunbo Zhang, Ke Huo, and Karthik Ramani: 'TRing: Instant and Customizable Interactions with Objects Using an Embedded Magnet and a Finger-Worn Device'. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. UIST '16. Tokyo, Japan: Association for Computing Machinery, 2016: pp. 169–181 (cit. on p. 72).

[Yu20] Yu, X., Z. Zhou, M. Xu, X. You, and X. Li: 'ThumbUp: Identification and Authentication by Smartwatch using Simple Hand Gestures'. *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. Los Alamitos, CA, USA: IEEE Computer Society, Mar. 2020: pp. 1–10 (cit. on p. 56).

[Yua08] Yuan, Bei and Eelke Folmer: 'Blind hero: enabling guitar hero for the visually impaired'. *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. 2008: pp. 169–176 (cit. on p. 27).

[Yuy15] Yuyang Ke, Yan Xiong, Yiqing Hu, Xudong Gong, and Wenchao Huang: 'Magemite: Character inputting system based on magnetic sensor'. *2015 IEEE 16th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. 2015: pp. 1–3 (cit. on p. 72).

[Zam07] Zamarripa, Myrna S., Victor M. Gonzalez, and Jesus Favela: 'The Augmented Patient Chart: Seamless Integration of Physical and Digital Artifacts for Hospital Work'. *Proceedings of the 4th International Conference on Universal Access in Human-computer Interaction: Applications and Services*. UAHCI'07. Beijing, China: Springer-Verlag, 2007: pp. 1006–1015 (cit. on p. 32).

[Zan18] Zang, Kaiyue, Jie Shen, Haosheng Huang, Mi Wan, and Jiafeng Shi: 'Assessing and Mapping of Road Surface Roughness based on GPS and Accelerometer Sensors on Bicycle-Mounted Smartphones'. en. *Sensors* (2018), vol. 18(3): p. 914 (cit. on p. 94).

[Zel03] Zelek, John S, Sam Bromley, Daniel Asmar, and David Thompson: 'A haptic glove as a tactile-vision sensory substitution for wayfinding'. *Journal of Visual Impairment & Blindness* (2003), vol. 97(10): pp. 621–632 (cit. on p. 27).

[ZEN] ZENRIN CO., LTD.: *Japanese Otaku City*. https://assetstore.unity.com/packages/3d/environments/urban/japanese-otaku-city-20359. Accessed: 2022-02-24 (cit. on p. 176).

[Zha03] Zhang, Bin, Sargur N. Srihari, and Sangjik Lee: 'Individuality of Handwritten Characters'. *Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2*. ICDAR '03. Washington, DC, USA: IEEE Computer Society, 2003: pp. 1086– (cit. on p. 39).

[Zha16a] Zhang, Cheng, A. Bedri, G. Reyes, Bailey Bercik, O. Inan, T. Starner, and Gregory D. Abowd: 'TapSkin: Recognizing On-Skin Input for Smartwatches'. *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces* (2016), vol. (cit. on p. 72).

[Zha17] Zhang, Cheng, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A. Cunefare, Omer T. Inan, and Gregory D. Abowd: 'SoundTrak: Continuous 3D Tracking of a Finger Using Active Acoustics'. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (June 2017), vol. 1(2) (cit. on pp. 25, 72).

[Zha16b] Zhang, L., F. Yang, Y. Daniel Zhang, and Y. J. Zhu: 'Road crack detection using deep convolutional neural network'. *2016 IEEE International Conference on Image Processing (ICIP)*. 2016: pp. 3708–3712 (cit. on p. 94).

[Zha21a] Zhang, Wenwen, Ralph Buehler, Andrea Broaddus, and Ted Sweeney: 'What type of infrastructures do e-scooter riders prefer? A route choice model'. *Transportation Research Part D: Transport and Environment* (2021), vol. 94: p. 102761 (cit. on p. 93).

[Zha18] Zhang, Yang and Chris Harrison: 'Pulp nonfiction: Low-cost touch tracking for paper'. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018: pp. 1–11 (cit. on p. 21).

[Zha15] Zhang, Yang and Chris Harrison: 'Quantifying the targeting performance benefit of electrostatic haptic feedback on touchscreens'. *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*. 2015: pp. 43–46 (cit. on p. 22).

[Zha19] Zhang, Yang, Wolf Kienzle, Yanjun Ma, Shiu S Ng, Hrvoje Benko, and Chris Harrison: 'ActiTouch: Robust touch detection for on-skin AR/VR interfaces'. *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 2019: pp. 1151–1159 (cit. on p. 25).

[Zha16c] Zhang, Yang, Junhan Zhou, Gierad Laput, and Chris Harrison: 'SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin'. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: Association for Computing Machinery, 2016: pp. 1491–1503 (cit. on pp. 25, 72).

[Zha21b] Zhao, Run, Dong Wang, Qian Zhang, Xueyi Jin, and Ke Liu: 'Smartphone-based Handwritten Signature Verification using Acoustic Signals'. *Proceedings of the ACM on Human-Computer Interaction* (2021), vol. 5(ISS): pp. 1–26 (cit. on p. 56).

[Zho16] Zhou, Junhan, Yang Zhang, Gierad Laput, and Chris Harrison: 'AuraSense: Enabling Expressive Around-Smartwatch Interactions with Electric Field Sensing'. Oct. 2016: pp. 81–86 (cit. on p. 72).

[Zim14] Zimmerman, Amanda, Ling Bai, and David Ginty: 'The gentle touch receptors of mammalian skin'. *Science (New York, N.Y.)* (Nov. 2014), vol. 346: pp. 950–4 (cit. on pp. 14, 16).

[Zim08] Zimmermann, Andreas and Andreas Lorenz: 'LISTEN: A User-Adaptive Audio-Augmented Museum Guide'. *User Modeling and User-Adapted Interaction* (2008), vol. 18(5): pp. 389–416 (cit. on pp. 155, 156).

[Zür19] Zürn, Jannik, Wolfram Burgard, and Abhinav Valada: 'Self-Supervised Visual Terrain Classification from Unsupervised Acoustic Feature Learning'. *arXiv preprint arXiv:1912.03227* (2019), vol. (cit. on pp. 94, 100).

List of Figures

List of Tables

Curriculum Vitae

Personal Data

Name	Maximilian Schrapel born 12.04.1989 in Magdeburg, Germany unmarried, german

Education

2016 - 2022	PhD Candidate Human-Computer Interaction Gottfried Wilhelm Leibniz Universität Hannover
2014 - 2016	Master of Science (M.Sc.) Technical Computer Science Gottfried Wilhelm Leibniz Universität Hannover Thesis title: 'Implementierung eines Frameworks zur Evaluation der videobasierten Objekterkennung mittels HOG und SVM in unterschiedlichen Szenarien'
2010 - 2014	Bachelor of Engineering (B.Eng.) Electrical Engineering Hochschule Hannover Thesis title: 'Machbarkeitsanalyse eines kopfgesteuerten Eingabeprinzips für angepasste Computertastaturen und -mäuse basierend auf einem Laser'
2009 - 2010	Fachhochschulreife, BBS Brinkstraße Osnabrück
2005	Realschulabschluss, Realschule Bad Iburg

Professional Experience

2015 - 2016	Tutor Gottfried Wilhelm Leibniz Universität Hannover
2014 - 2015	Working Student arcutronix GmbH
2010 - 2014	Working Student Resch Electronic Innovation GmbH

2009	Technical Quality Assurance Associate Resch Electronic Innovation GmbH
2005 - 2009	Apprenticeshipment Electronics Technician for Devices and Systems Resch Electronic Innovation GmbH

Complete List of Publications

Journal Paper

- SCHRAPEL, MAXIMILIAN AND HAPPE, JANKO AND ROHS, MICHAEL: 'EnvironZen: Immersive Soundscapes via Augmented Footstep Sounds in Urban Areas'. *i-com* (2022), vol. 21(2): pp. 219–237

Full Papers

- SCHRAPEL, MAXIMILIAN & STADLER, MAX-LUDWIG & ROHS, MICHAEL: 'Pentelligence: Combining Pen Tip Motion and Writing Sounds for Handwritten Digit Recognition'. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2018: pp. 1–11
- SCHRAPEL, MAXIMILIAN & SCHULZ, THILO & ROHS, MICHAEL: 'Augmenting Public Bookcases to Support Book Sharing'. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. MobileHCI '20. Oldenburg, Germany: Association for Computing Machinery, 2020
- SCHRAPEL, MAXIMILIAN et al.: 'Sign H3re: Symbol and X-Mark Writer Identification Using Audio and Motion Data from a Digital Pen'. *Proceedings of Mensch Und Computer 2022*. MuC '22. Darmstadt, Germany: Association for Computing Machinery, 2022: pp. 209–218
- SCHNEIDER, KURT AND BUSCH, MELANIE AND KARRAS, OLIVER AND SCHRAPEL, MAXIMILIAN AND ROHS, MICHAEL: 'Refining vision videos'. *International Working Conference on Requirements Engineering: Foundation for Software Quality*. Springer. 2019: pp. 135–150
- KAUL, OLIVER BEREN AND DOMIN, ANDREAS AND ROHS, MICHAEL AND SIMON, BENJAMIN AND SCHRAPEL, MAXIMILIAN: 'VRTactileDraw: A Virtual Reality Tactile Pattern Designer for Complex Spatial Arrangements of Actuators'. *IFIP Conference on Human-Computer Interaction*. Springer. 2021: pp. 212–233

Short Papers & Posters

- SCHRAPEL, MAXIMILIAN & HERZOG, FLORIAN & RYLL, STEFFEN & ROHS, MICHAEL: 'Watch My Painting: The Back of the Hand as a Drawing Space for Smartwatches'. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery, 2020: pp. 1–10

- Schrapel, Maximilian & Liebers, Jonathan & Rohs, Michael & Schneegass, Stefan: 'Skiables: Towards a Wearable System Mounted on a Ski Boot for Measuring Slope Conditions'. *19th International Conference on Mobile and Ubiquitous Multimedia*. MUM 2020. Essen, Germany: Association for Computing Machinery, 2020: pp. 320–322
- Schrapel, Maximilian & Etgeton, Philipp & Rohs, Michael: 'SpectroPhone: Enabling Material Surface Sensing with Rear Camera and Flashlight LEDs'. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery, 2021
- Pohl, Henning and Becke, Dennis and Wagner, Eugen and Schrapel, Maximilian and Rohs, Michael: 'Wrist compression feedback by pneumatic actuation'. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 2015: pp. 315–318
- Chazette, Larissa et al.: 'Designing Software Transparency: A Multidisciplinary Endeavor'. *Softwaretechnik-Trends* (2020), vol. 40(1): pp. 5–6

Workshops

- Maximilian Schrapel & Anne Finger & Jochen Meyer & Michael Rohs & Johannes Schöning & Alexandra Voit: *Ubiactivity'18: International Workshop on Integrating Physical Activity and Health Aspects in Everyday Mobility. UbiComp'18, Singapore*. 2018
- Shadan Sadeghian & Maximilian Schrapel et al.: *Workshop on Designing Technologies for Future Forms of Sustainable Mobility. MobileHCI'20, Oldenburg, Germany*. 2020

Workshop Paper

- Maximilian Schrapel, Anne Finger and Michael Rohs: 'Integrating Recommended Physical Activity in Everyday Mobility'. *Accepted workshop papers at the workshop on Augmented Humanity using Wearable and Mobile Devices for Health and Wellbeing at MobileHCI'18* (2018), vol.

Other Articles

- Anne Finger and Lena Greinke and Maximilian Schrapel: 'Fußverkehr als Beitrag zur Gesunden Stadt'. *PLANERIN 5/2018: Gesundheitsversorgung - Gesund leben in Stadt, Land und Quartier* (2018), vol. 5

- GuS Kommunikation and Maximilian Schrapel: 'MINT-Macher im Gespräch: Maximilian Schrapel'. *TTS: Treffpunkt Technik in der Schule* (2021), vol. 2

Patent

- Maximilian Schrapel, Luise aus der Fünten (EZN): *Pentelligence: Handschrifterkennung mittels Audio und Bewegungsdaten von Stiften.* DE Patent 10 2018 107 409 A1. 2018

Awards

- Maximilian Schrapel, Thilo Schulz, and Michael Rohs. *MobileHCI 2020 Best Paper Award* [Sch20c]
- Maximilian Schrapel, Dennis Grannemann, and Michael Rohs. *MuC 2022 Honorable Mention Award* [Sch22a]
- Maximilian Schrapel, Sergej Löwen, and Michael Rohs. InnovateFPGA Design Contest 2019: European Final Bronze Award, *MultiWave: A Fully Customizable Mobile Function Generator for Haptic Feedback in VR*
- Maximilian Schrapel and Michael Rohs. *starting business: Science Award.* 2021

Colophon

This thesis was typeset with Overleaf LaTeXeditor. The template Version 3.2.5 developed by Matthias Pospiech was used. The Figures were created with Seaborn and Inkscape.