Auditory Displays for People with Visual Impairments during Travel

Zur Erlangung des akademischen Grades einer
Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

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

Dissertation

von

Angela Schön
geborene Constantinescu
aus Galați, Rumänien

Tag der mündlichen Prüfung:	21.11.2023
Hauptreferent:	Prof. Dr.-Ing. Rainer Stiefelhagen
Korreferent:	Prof. Dr. rer. nat. habil. Gerhard Weber

Kurzfassung

Menschen mit Blindheit oder Sehbehinderungen begegnen beim Reisen zahlreichen Barrieren, was sich auf die Lebensqualität auswirkt. Obwohl spezielle elektronische Reisehilfen schon seit vielen Jahren im Mittelpunkt der Forschung stehen, werden sie von der Zielgruppe nach wie vor kaum genutzt. Dies liegt unter anderem daran, dass die von den Nutzern benötigten Informationen von der Technologie nur unzureichend bereitgestellt werden. Außerdem entsprechen die Schnittstellen selten den Bedürfnissen der Nutzer. In der vorliegender Arbeit gehen wir auf diese Defizite ein und definieren die Anforderungen für barrierefreies Reisen in Bezug auf den Informationsbedarf (Was muss vermittelt werden?) und die nichtfunktionalen Anforderungen (Wie muss es vermittelt werden?). Außerdem schlagen wir verschiedene auditive Displays vor, die die Bedürfnisse von Menschen mit Sehbeeinträchtigungen während einer Reise berücksichtigen. Wir entwerfen, implementieren und evaluieren unsere Schnittstellen nach einem nutzerzentriertem Ansatz, wobei wir während des gesamten Prozesses Nutzer und Experten aus diesem Bereich einbeziehen.

In einem ersten Schritt erheben wir den Informationsbedarf von Menschen mit Behinderungen im Allgemeinen und von Menschen mit Sehbeeinträchtigungen im Besonderen, wenn sie sich in Gebäuden bewegen. Außerdem vergleichen wir die gesammelten Informationen mit dem, was derzeit in OpenStreetMap (OSM), einer freien geografischen Datenbank, kartiert werden kann, und machen Vorschläge zur Schließung der Lücke. Unser Ziel ist es, die Kartierung aller benötigten Informationen zu ermöglichen, um sie in Lösungen zur Unterstützung des unabhängigen Reisens zu verwenden.

Nachdem wir die Frage beantwortet haben, welche Informationen benötigt werden, gehen wir weiter und beantworten die Frage, wie diese den Nutzern vermittelt werden können. Wir definieren eine Sammlung nicht-funktionaler Anforderungen, die wir in einer Befragung mit 22 Mobilitätstrainern verfeinern und bewerten.

Anschließend schlagen wir eine Grammatik - oder anders ausgedrückt, eine strukturierte Art der Informationsvermittlung - für Navigationsanweisungen bei Reisen im Freien vor, die Straßenränder, das Vorhandensein von Gehwegen und Kreuzungen berücksichtigt - alles wichtige Informationen für Menschen mit Sehbeeinträchtigungen. Darüber hinaus können mit unserer Grammatik auch Orientierungspunkte, Sehenswürdigkeiten und Hindernisse vermittelt werden, was die Reise zu einem ganzheitlichen und sichereren Erlebnis macht. Wir implementieren unsere Grammatik in einen bestehenden Prototyp und evaluieren sie mit der Zielgruppe.

Es hat sich gezeigt, dass in Gebäuden Beschreibungen der Umgebung die Erstellung von mentalen Karten unterstützen und damit die Erkundung und spontane Entscheidungsfindung besser fördern als Navigationsanweisungen. Wir definieren daher eine Grammatik für die Vermittlung von Informationen über die Umgebung in Innenräumen für Menschen mit Sehbeeinträchtigungen. Wir bewerten die

Grammatik in einer Online-Studie mit 8 Nutzern aus der Zielgruppe. Wir zeigen, dass die Nutzer strukturierte Sätze mit fester Wortreihenfolge benötigen. Schließlich implementieren wir die Grammatik als Proof-of-Concept in eine bestehende prototypische App.

Sprachausgabe ist zwar Stand der Technik im Bereich der Ausgabeschnittstellen für Menschen mit Sehbeeinträchtigungen, hat aber auch Nachteile: es ist für Menschen mit Leseschwäche unzugänglich und kann für manche Nutzer zu langsam sein. Wir nehmen uns dieses Problems an und untersuchen den Einsatz von Sonifikation in Form von auditiven Symbolen in Kombination mit Parameter-Mapping zur Vermittlung von Informationen über Objekte und deren Verortung in der Umgebung. Da eine erste Evaluierung positive Ergebnisse lieferte, erstellten wir in einem nutzerzentrierten Entwicklungsansatz einen Datensatz mit kurzen auditiven Symbolen für 40 Alltagsgegenstände. Wir evaluieren den Datensatz mit 16 blinden Menschen und zeigen, dass die Töne intuitiv sind. Schließlich vergleichen wir in einer Nutzerstudie mit 5 Teilnehmern Sprachausgabe mit nicht-sprachlicher Sonifikation. Wir zeigen, dass Sonifikation für die Vermittlung von groben Informationen über Objekte in der Umgebung genau so gut geeignet ist wie Sprache, was die Benutzerfreundlichkeit angeht. Abschließend listen wir einige Vorteile von Sprache und Sonifikation auf, die zum Vergleich und als Entscheidungshilfe dienen sollen.

Diese Arbeit befasst sich mit den Bedürfnissen von Menschen mit Sehbeeinträchtigungen während der Reise in Bezug auf die benötigten Informationen und Schnittstellen. In einem nutzerzentrierten Ansatz schlagen wir verschiedene akustische Schnittstellen vor, die auf sprachlicher und nicht-sprachlicher Sonifikation basieren. Anhand mehrerer Nutzerstudien, an denen sowohl Nutzer als auch Experten beteiligt sind, entwerfen, implementieren und evaluieren wir unsere Schnittstellen. Wir zeigen, dass elektronische Reisehilfen in der Lage sein müssen, große Mengen an Informationen auf strukturierte Weise zu vermitteln, jedoch angepasst an den Nutzungskontext und die Präferenzen und Fähigkeiten der Nutzer.

Abstract

People with visual impairments face multiple barriers while traveling, with consequences on the quality of life. While specific electronic travel aids have been in the focus of research for many years now, their adoption by the target user group remains scarce. Reasons include the fact that information needed by the users is insufficiently provided by technology. Moreover, the interfaces rarely suit the needs of their users. We address these shortcomings and define requirements for accessible travel in terms of informational needs (*what* to convey) and non-functional requirements (*how* to convey it). Moreover, we propose several auditory displays that consider the needs of people with visual impairments while travelling. We design, implement and evaluate our interfaces in a user-centered approach, involving many users and experts in the field throughout the process.

In a first step, we gather the informational needs of people with disabilities in general and people with visual impairments in particular, when travelling inside buildings. Moreover, we compare the information gathered with what can be currently mapped in OpenStreetMap (OSM), a free geographic database, and make suggestions for closing the gap. Our goal is to enable the mapping of all the information needed, to be used in solutions that support independent travel.

After having answered the question of *what* information is needed, we go on and answer the question of *how* to convey it to the users. We define a set of non-functional requirements which we refine and assess in an expert survey with 22 mobility trainers.

We then propose a grammar - or, in other words, a structured way of conveying information - for navigation instructions during *outdoor* travel that takes into account roadsides, the existence of sidewalks and crossings - all relevant information for people with visual impairments. Moreover, landmarks, points of interest and obstacles can also be conveyed with our grammar, making the travel a holistic and safer experience. We implement our grammar into an existing prototype and evaluate it with the target user group.

It has been shown that indoors, descriptions of the environment support the creation of mental maps, and with that, foster exploration and spontaneous decision making, better than navigation instructions. We thus define a grammar for conveying *indoor* information about the environment to people with visual impairments. We assess the grammar in an online study with users from the target group. We show that users need structured sentences with fixed word order. Finally, we implement the grammar, as a proof-of-concept, into an existing prototypical app.

While text-to-speech is state-of-the-art in output interfaces for people with visual impairments, it also has disadvantages: it is inaccessible to illiterate people and can be too slow for some users. Thus, we address this issue and investigate the use of sonification in form of short auditory icons combined with parameter mapping for conveying information about objects in the environment. Since a preliminary evaluation yielded positive results, we created, in a user-centered development approach, a dataset of *short* auditory icons for 40 everyday objects. We evaluate the dataset with 16 blind people and show that the sounds are intuitive. Finally, we compare text-to-speech with non-speech sonification in a user study with 5 participants. We show that for conveying rough information about objects and their location in the environment, sonification is as suitable as speech in terms of usability. We conclude by listing some of the advantages of both speech and sonification, to serve as comparison and as a help in decision-making.

This thesis addresses the needs of people with visual impairments during travel in terms of information required and interfaces. In a user-centered approach we propose several auditory displays based on speech and non-speech sonification. Through several user studies involving both users and experts, we design, implement and evaluate our interfaces. We show that electronic travel aids must be able to convey large amounts of information in a structured way, but adapted to the context of use and user's preferences and abilities.

Acknowledgements

I would like to thank first of all Prof. Rainer Stiefelhagen, for providing me throughout all the years with the possibility to work in his teams, to write this thesis and for his most valuable feedback. I also thank Prof. Gerhard Weber for co-supervising my thesis.

This thesis would not have been possible without the support of Dr. Karin Müller. She always gave me the right impulses at the right time, she encouraged me and managed projects in such a way so as to allow me to carry the research that led to this thesis. I owe this to you!

I would also like to thank the wonderful team at ACCESS@KIT (former SZS) who have always been there for me: Angelika, Ann-Christin, Christina, Gerhard, Giuseppe, Julia, Susanne, Thorsten and Vanessa. A big thank you especially to Gerhard, who had the patience to always try my ideas and give me most valuable feedback. I also thank the CV:HCI team, which I was a part of for a short time, and with whom I collaborated for the rest of my research. I am grateful to you all, but will only mention here: Alina, Daniel, Corinna, Jiaming, Kailun, Manel and Monica. Thank you for supporting me!

Finally, I would like to thank my family - I would not have made it without your encouragement and support: my husband, my children and most importantly my parents, who took over a great deal of my responsibilities at home so I can do this. I am blessed to have you all.

Glory to God for All Things!

Contents

Kurzfassung . . . . **i**

Abstract . . . . **iii**

Acknowledgements . . . . **v**

Acronyms and Definitions . . . . **ix**

1. Introduction **1**
  1.1. Motivation 1
  1.2. Thesis Road Map and Contributions 4
  1.3. Published Contributions and Credits 6

2. Background and Related Work **7**
  2.1. The Needs of People with Visual Impairments During Travel 7
  2.2. Output Interfaces for People with Visual Impairments 11
    2.2.1. Speech Auditory Displays 12
    2.2.2. Non-Speech Auditory Displays 13

3. Requirements for Accessible Travel **19**
  3.1. Indoor Accessibility Features 19
    3.1.1. Literature Review on Accessibility Features 20
    3.1.2. Accessibility Features for People with Visual Impairments 28
    3.1.3. Conclusion 30
  3.2. Non-Functional Requirements for Assistive Output Interfaces 31
    3.2.1. Initial Set of Requirements 32
    3.2.2. Expert Assessment 33
    3.2.3. Refinement of Requirements 36
    3.2.4. Conclusion 37

4. Speech Auditory Displays **39**
  4.1. Speech Auditory Display for Outdoor Travel: Enhanced Navigation Information and Objects Announcements 39
    4.1.1. Creation of the Grammar through an Expert Round-Table 40
    4.1.2. Prototypical Implementation of the Outdoor Grammar 45
    4.1.3. Evaluation 49
    4.1.4. Comparison with State-of-the-Art 58
    4.1.5. Conclusion 60

4.2. Speech Auditory Display for Indoor Travel: Grammar for POI-based Descriptions of the Surroundings . . . 61
4.2.1. Analysis of Literature and Existing Solutions . . . 62
4.2.2. Grammar Specification . . . 64
4.2.3. Evaluation of the Indoor Grammar . . . 65
4.2.4. Proof-of-Concept: Prototypical Implementation of the Indoor Grammar . . . 73
4.2.5. Conclusion . . . 75

5. Non-Speech Auditory Displays . . . 77
5.1. Preliminary Study: Outdoor Object Sonification . . . 77
5.2. Creation and Evaluation of Short Auditory Icons for People with Visual Impairments . . . 90
5.2.1. Assessment of Conceptual Mappings through Online User Study . . . 90
5.2.2. Creation of Short Auditory Icons Set for Everyday Objects . . . 95
5.2.3. Evaluation of the Auditory Icons Dataset . . . 98
5.2.4. Discussion . . . 102
5.2.5. Conclusion . . . 104
5.3. Comparison Between Speech and Non-Speech Sonification . . . 105

6. Conclusion . . . 113
6.1. Contributions . . . 113
6.2. Future Work . . . 118

A. Appendix . . . 119
A.1. Complete Outdoor Grammar . . . 119
A.2. Thematic Analysis of Conceptual Mappings . . . 120

List of Publications . . . 129

Bibliography . . . 131

Acronyms and Definitions

ACCESS@KIT	Center for Digital Accessibility and Assistive Technology at KIT
Accessibility feature	We define as accessibility feature any object or thing that influences, either positively or negatively, people with disabilities during orientation and independent travel. Accessibility features include landmarks, points of interest and obstacles
Auditory display	An output interface consisting only of sound - whether speech or non-speech
Bone conducting headphones	Mainstream headphones that are placed on the bone in front of the ear. The sound is transmitted through the skull's bones, leaving the ear open for environmental sounds. They are commercially available and used for instance by sighted cyclists, joggers as well as by people with visual impairments
ETA	Electronic Travel Aid. Any mobile application or device that supports people with visual impairments during travel. ETAs can be used, for instance, for obstacle avoidance or to provide navigation instructions
KIT	Karlsruhe Institute of Technology
Landmark	Landmarks are "features that can provide physical or tactile cues to help users to confirm their location such as tactile paving, doorways, and other building infrastructures (e.g. elevator, escalator, and stairs)" (Sato et al. 2017)
ms	Milliseconds
NASA – TLX	Nasa Task Load Index. Questionnaire designed by NASA to assess the cognitive load of a person during or after carrying out a task
NPS	Net Promoter Score. A single-question questionnaire meant to estimate the likelihood that users promote a product
OSM	OpenStreetMap. An open and free geographic database, whose content is added collaboratively by volunteers and whose data can be used by everyone. In short, a free map of the world. Recently, it added indoor features so one can also map buildings

POI	Point of Interest. “POIs, such as shops and facilities (e.g., restroom), are defined as places that might interest users during navigation” (Sato et al. 2017)
Raw TLX	or RTLX is a simplified variation of the NASA-TLX questionnaire, in which the weighting is dropped
Screen reader	A program that “reads out” the contents on a screen, organizes them and conveys them via text-to-speech or renders them on a Braille display. Screen readers enable blind people to use computers and smartphones
SD	Standard deviation
SZS	Study Center for Visually Impaired Students - the name of ACCESS@KIT before 2022
VI	Visual Impairment. In Germany, visual impairment is defined as visual acuity of less than 0.3, while blindness is a visual acuity of less than 0.02 for both eyes, according to Part A.6.a of the Annex to § 2 VersMedVO (Bundesministerium der Justiz 2008). “Visual acuity is defined as the ability to read a standard test pattern at a certain distance, usually measured in terms of a ratio to ’normal’ vision” (John M Evans LLC 2006)

1. Introduction

1.1. Motivation

In 2010 there were an estimated 285 million people with visual impairments in the world (Hersh 2018), while more current reports of the World Health Organization name as many as 2.2 billion[1]. In Germany alone live approximately 1.2 million of them[2] - almost 335 thousands of whom are *severely* visually impaired according to a 2021 report of the Federal Statistical Office (Böhm 2021). Independent travel is one of the major challenges faced by people with visual impairments, which in turn affects their quality of life (Goldschmidt 2018).

When travelling through *known* environments, people with visual impairments use the knowledge learned in the orientation and mobility training to reach their destination safely. But high traffic, open spaces, wide streets to be crossed, crowds of people, loud environments, movable objects such as parked bicycles or chairs and unforeseen changes such as construction sites can still pose great dangers. In *unknown* environments, the situation is even more challenging. When travelling in unknown buildings, for instance, the most popular strategy is to ask sighted passers-by for directions - if there are any; but visually impaired people wished they had more independence, for instance by using maps or descriptions (Müller et al. 2022). In a research study that interviewed 20 visually impaired individuals in the US and South Korea about travelling to unknown destinations, the participants also said they sometimes rely on directions from sighted people, but mentioned that these are often inadequate, and they would do this only “if they felt particularly brave” (Quinones et al. 2011). Even when travelling with a sighted guide, one often gets unreliable directions due to sighted people not fully understanding the needs of the visually impaired people: “despite the advantages of having sighted guides, several works have observed unreliable guidance by sighted people due to their lack of knowledge about how visually impaired people navigate or verbal descriptions methods to give environmental cues [...]. Navigators have proclaimed their experiences of receiving irrelevant information, ambiguous phrases like ‘there’, or inaccurate measurement estimations” (Kamikubo et al. 2020). One solution is to use assistive technologies to get the information needed for independent travel.

With recent advances in geopositioning, solutions such as pedestrian navigation apps have gained in popularity. But, while for the most of us, tools like Google Maps have become indispensable, for people with visual impairments these are not enough. Even solutions specifically designed for this user group, the so-called electronic travel aids, “have proven too generalist to be useful” (Nicolau et al. 2009b) and “fail to address the real needs of their users” (Dramas et al. 2008) - and this is as

---

[1] `https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment`

[2] `https://www.rehadat-statistik.de/statistiken/behinderung/behinderungsarten/blindheit-und-sehbehinderung`

true today as it was 15 years ago, when these statements were made. This results in low usage of mobility devices (Hersh 2018), except for the white cane. Besides very precise navigation instructions - more precise than mainstream technologies provide nowadays - people with visual impairments also need information about *landmarks* along the route (Gomez et al. 2016), *points of interest* (POI) and *obstacles*. For instance, obstacles at the ground level and about 1.5 meters in front of the user are mainly covered by the swipe of the white cane. But everything beyond that, and especially objects at head level which can cause serious injuries and objects beyond 1.5 meters are mostly inaccessible. Due to the swiping movement of the cane combined with the fast walking speed of some blind people, even short range objects at ground level can be missed, such as bicycles, when the cane swipes just in between the wheels, causing the user to bump into the handlebars. Landmarks, on the other hand, are objects used as reference points to ease the orientation and way finding. Sighted people usually orient using visual landmarks, while visually impaired people use different kinds of landmarks, from acoustic ones (such as buses stopping at a bus stop, doors opening or closing, birds singing, voices of children playing on a playground, the reverberation of a large lobby or of a narrow corridor, phone or keyboard sounds) to tactile (the ground texture and changes thereof under the feet, grass on the side of the road felt with the white cane, material and structure of doors felt with the hands), olfactory landmarks (the smell from the bakery, exhaust emissions at a bus stop) and also visual, for the ones who have some remaining sight (such as light contrast, sources of light including light bulbs and windows, strong color contrasts). Only little of this information is provided by electronic travel aids, and as we have already stated, sighted people "offer instructions based on their visual capabilities, which are insufficient for the visually impaired [people]" (Gomez et al. 2016). Thus, providing information about *non-visual* landmarks to people with visual impairments within navigation instructions helps orient and stay on route. Information about *visual* landmarks, on the other hand, helps to better understand directions obtained from sighted people. Or, when combined with navigation instructions that address their needs, it can help them better navigate unknown environments by setting visual landmarks the way sighted people do.

Advances in geopositioning and deep learning could soon provide all the missing information that people with visual impairments need. But while technology is making great leaps forward, the issue of designing interfaces to meet the users' needs remains inadequately addressed. Especially for people with visual impairments, "feedback is essential in navigation tasks [...]. A wrong choice may result in hazardous and potentially accident-prone situations for the user" (Walle et al. 2022). Imagine a blind person travelling through a city, listening to the environment for walking and crossing safely, while an assistive system starts giving speech-based information when the user is in the middle of the traffic lane, covering or obstructing the more important traffic sounds, and posing a serious safety threat. The output of mainstream and frequently even the one of assistive systems is often taken for granted, while the development focuses on the technology and less on the interface. This so-called technology-driven development is one of the reasons why, despite many years of research in assistive technologies, the white cane is still the travel aid most used by blind people throughout the world today.

For people with visual impairments, and especially for blind people, sound is the most popular (Walle et al. 2022) and convenient alternative to vision for perceiving the environment. The "use of sound to understand the world, [...] the technique of rendering sound in response to data and interactions" is called sonification (Hermann et al. 2011). Sonification encompasses speech and non-speech techniques. Speech is the sonification most widely used by blind people nowadays. Non-speech sonification, on the other hand, is used very scarcely in mainstream apps and systems, mostly as short sounds for announcing events or as alarms. An example of a popular non-speech sonification is the car parking system that beeps faster the closer the car gets to an obstacle. Sonification, along with all other aspects of a human-computer interaction system, such as the setup, peripherals, modes of interaction, etc., is referred to as "auditory display" (Hermann et al. 2011). Although sonification encompasses both text-to-speech and non-speech techniques, in this thesis we will sometimes refer to sonification to describe non-speech techniques alone, as it is currently the practice in this community[3].

Besides sonification, there are other interfaces appropriate for people with visual impairments. The haptic ones, such as Braille and pin matrix displays, vibrating armbands or simply the vibrating motor of a smart watch or a smartphone, and even force feedback, are among the most practical, after audio interfaces. Less common interfaces include the ones based on smell, taste (Kortum 2008) or even heat. Most of these alternative interfaces require bulky or expensive hardware, are not mobile, lack intuitiveness or have limited application. The vibrotactile feedback of a smartphone, for instance, cannot encode much information intuitively using just one motor. So until mainstream products will feature truly mobile alternative interfaces and output devices, auditory displays remain the most versatile, cheap and practical option which can convey large amounts of information.

But even in the case of auditory displays, a lot remains to be done for them to be truly useful to people with visual impairments during travel. Speech, for instance, is taken for granted by developers as being intuitive. Little thought is put into the design of the speech instructions. Experts in orientation and mobility (O&M) for people with visual impairments, namely O&M trainers, noticed, however, that each user has different language skills and vocabulary. This is why they adapt the contents and especially the language of their coaching to each user. Randomly chosen and poorly designed speech instructions will, at best, scare off potential users, and in the worst case, can pose a danger to their users. Moreover, people with visual impairments need a large amount of information, so a balance must be found between conveying too much and not enough information. What information precisely is needed is another aspect that needs to be addressed. While outdoors this is better known due to the existing solutions, indoors it is less clear what people with visual impairments need to know about when travelling. Finally, alternative auditory displays, such as non-speech sonification, might prove more effective than speech in conveying some of the information needed. A danger, for instance, is better announced through a very short, stringent tone, than through text-to-speech. Auditory displays have only been used sparsely in digital solutions, as the focus has always been on graphical interfaces.

---

[3] The Sonification Handbook (Hermann et al. 2011), for instance, does not name text-to-speech among the sonification techniques, although the International Community for Auditory Display (ICAD) recognizes that, according to the definition of sonification, this is indeed the case. Some view text-to-speech as a type of parametric mapping, which is one of the currently accepted sonification techniques (see `http://icad.org/pipermail/community_icad.org/2023-January/000006.html`). Since auditory display is a little explored and not so well established research field, definitions and classifications are subject to change. For the purpose of this thesis, we need to differentiate between speech and non-speech auditory displays, so we will at times call the latter generically *sonification*.

Recently, speech-based interfaces, mostly known as voice user interfaces, have gained in popularity. Sonification, on the other hand, remains a hardly explored subject.

1.2. Thesis Road Map and Contributions

In this thesis, we take a *user-centered approach* to address two main questions related to people with visual impairments in the context of travel:

1. *What* information is relevant for people with visual impairments during travel?, and
2. *How* should this information be conveyed effectively through speech and sound?

Through several user studies and with the constant involvement of users and experts in all phases of the development, starting with the requirements gathering, during implementation and in particular during evaluation, we designed speech and non-speech auditory displays for people with visual impairments during indoor and outdoor travel.

In the following paragraphs we give an overview of the structure of this thesis and outline the main contributions made.

Chapter 2. Background and Related Work. In Chapter 2, Section 2.1 we start by giving an overview of the needs of people with visual impairments during travel, as accounted for in the literature. Besides listing some of the problems faced by the user group, we highlight studies that deal with the *information needs* of people with visual impairments and also give a short overview of *electronic travel aids*. In Section 2.2 we then give a quick overview of some of the output interfaces that are appropriate for people with visual impairments, especially in the context of travel. We argue why auditory displays are currently the most feasible option and list some of the most prominent works that deal with either speech or non-speech auditory displays. We give here an introduction into the topic of *sonification*, as it is a little known area. We list the currently accepted sonification techniques and go into slightly more detail into the topic of *auditory icons* and their conceptual mappings. We end the chapter with a short review of sonification research for people with visual impairments in the travel context and of object sonification.

Chapter 3. Requirements for Accessible Travel. In Chapter 3, Section 3.1, we start by gathering from literature the *information need* of people with disabilities in general, and of people with visual impairments in particular, when travelling indoors - this is what we call indoor accessibility features. We gather this data into a database, together with the information that can be mapped into OpenStreetMap - short OSM, a free and open geographic database. We compare the two sources in an alignment table of our database in order to uncover gaps and to make proposals that would allow the mapping of more accessibility features into the free OSM world map. Finally, through a collaboration with Sozialhelden e.V., we set the cornerstone for the standardization of the accessibility features.

After assessing *what* information is needed, we show in Section 3.2 *how* this information should be conveyed to the user. Through literature review and a survey with 22 experts in orientation and mobility (O&M) for people with visual impairments, we define a set of *13 non-functional requirements* for the output interface of mobile assistive systems for people with visual impairments. We also provide ratings to these requirements based on the experts' assessment, showing that all requirements stated are important. We finally compile a set of sample criteria for each requirement. Our requirements list can be used during the design specification or for the heuristic evaluation of auditory displays for people with visual impairments.

Chapter 4. Speech Auditory Displays. Having defined what information is needed and how it should be conveyed, we go on and propose, in Chapter 4, speech-based auditory displays for people with visual impairments in the context of travel. In Section 4.1, we propose a grammar for *outdoor* navigation instructions enhanced with information about objects (accessibility features) in the environment. The grammar is refined in a focus group with 5 experts in O&M and accessibility. We integrate the grammar into a prototypical system that gives route instructions tailored for this target user group while at the same time recognizes crossing elements such as traffic lights and zebra crossings together with their exact location with respect to the user. The speech instructions and the information about the objects are smoothly integrated, resulting into one consistent interface. We evaluate the prototype in an on-site user study with 15 people with visual impairments in a city. We discuss the results of the evaluation and present the improvements suggested by the participants.

In the next Section, 4.2, we conduct an analysis of existing literature to define a grammar for *indoor* use. Instead of focusing on navigation, we choose to convey only information about the accessibility features around the user, giving thus an overview of the environment. This fosters independence by helping users to orient themselves, find things and discover interesting locations. We evaluate egocentric textual descriptions generated with the grammar in an online user study with 8 blind people. We show that users could create a mental map of the environment only by reading the instructions provided. We also found that users wish to have a structured grammar with fixed word order, and that a random word order within sentences frustrates users, resulting in a higher cognitive load. Users also provided improvement suggestions for the grammar and any system implementing it. Finally, taking the user feedback into account, we implemented the grammar into an existing prototypical Flutter app, as a proof-of-concept.

Chapter 5. Non-Speech Auditory Displays. Well-designed speech interfaces help users form mental maps of their environment and navigate more confidently, as our evaluations in Chapter 4 show. But speech also has its limitations. For instance, for some users or in certain situations, it is too slow. Besides, the information that needs to be conveyed is very large, as our lists in Section 3.1 shows, and conveying all of it with one modality - speech - can result in cognitive overload. Thus, in Chapter 5, we wanted to know whether sonification can successfully replace speech in given situations. We start in Section 5.1 with a preliminary user study in order to assess whether *short* auditory icons combined with parameter mapping are suitable for conveying information about objects in the environment to the user. In a qualitative on-site evaluation with 5 people with visual impairments, we showed that this is the case. Moreover, we showed that the used auditory icons were on average intuitive and, in any case, they could be learned very fast. The study also suggested that conceptual mappings should be further investigated. Thus, in Section 5.2 we start by assessing the conceptual mappings for a set of everyday objects, or, in other words, the sounds associated with the given objects. We do this in an online open-text questionnaire with 11 people with visual impairments. Based on these answers we created, with the involvement of a blind person, 40 short auditory icons in two versions: 500 ms and 1000 ms ones. We finally evaluated the 500 ms auditory icons in another online questionnaire with 16 blind people. The results of the evaluation as well as the suggestions for improvement are discussed at the end of this section.

Finally, in Section 5.3 we wanted to know if sonification can, in given scenarios, replace speech in terms of usability. For speech, we used a reduced version of the grammar defined in Subsection 4.2.2 that conveyed the exact same information as the auditory icons with parameter mapping - which is the same sonification used in Section 5.1. This time, however, we didn't only sonify the environment in front of the user, but 360° around. We called a small expert round table in order to establish a suitable sonification method, as 3D sound behind the user cannot be conveyed on headphones appropriate for people with visual impairments, namely bone conducting headphones. In a subsequent comparison evaluation with 5 users inside a building, we assessed the effectiveness, efficiency and user satisfaction of both modalities - speech and sonification. The study showed that for getting an overview of the environment, without many details, sonification is at least as good as text-to-speech. The participants also proposed several contexts in which they would like to use sonification displays based on auditory icons and parameter mapping.

1.3. Published Contributions and Credits

Chapter 2, Section 3.1 and Subsections 4.1.1, 4.2 and 5.1 are partially based on our publications Constantinescu et al. (2022a, 2019, 2022b, 2020) and Engel et al. (2020). Angela Constantinescu has put much effort in ensuring that all the material in this thesis includes only her own work and ideas. Only relevant parts of the publications, moreover only those that constitute the work and ideas of Angela Constantinescu have been included in this thesis. While the concerned text from the publications was in most cases supplemented with additional information, there can be situations in which text from the mentioned publications was reproduced ad litteram in this thesis. For readability reasons, quotations have been omitted in these cases.

Parts of Chapter 2 are based on all mentioned publications. Section 3.1 is based on our ICCHP publication Constantinescu et al. (2022a) © Springer. Subsection 4.1.1 is based on our ASSETS publication Constantinescu et al. (2019). Section 4.2 is based on our ICCHP publication Constantinescu et al. (2022b) © Springer. Section 5.1 is based on our ICMI publication Constantinescu et al. (2020). Sections 4.2 and 5.2 resulted from the joint research conducted by the master student Eva-Maria Neumann (Neumann 2022), the bachelor student Lukas Fritsch (Fritsch 2020) and Angela Constantinescu and Dr. Karin Müller as their supervisors. In the case of both bachelor and master theses, Angela Constantinescu was responsible for providing the ideas (together with Dr. Karin Müller, who co-supervised the theses), and particularly for interpreting the results of the evaluations performed. The students Eva-Maria Neumann and Lukas Fritsch were mainly responsible for the practical part (implementation, creation of the sounds).

Please note that for clarity reasons and without loss of generality, this document is not gendered.

2. Background and Related Work

While there are a few research efforts in the area of travel assistance for people with visual impairments, not many of them focus or even consider the interface. In this thesis, we address auditory displays for people with visual impairments during travel. We propose both speech-based and sonification approaches for conveying information about the environment and about the route to people with visual impairments during outdoor or indoor travel. This chapter provides an overview of relevant background information and related work. We start by giving an overview of the needs of people with visual impairments during travel. This includes a quick review of both the informational needs and the electronic travel aids used by people with visual impairments. Afterwards, we take a look at the various interfaces that can be used for this target group and argue why we focus on auditory ones. Finally, we give an overview on auditory displays, both speech-based and sonification ones. We insist more on sonification and give an introduction to this topic, as it is a little explored and, although not new - a hardly known field.

2.1. The Needs of People with Visual Impairments During Travel

There are various research projects that deal with the orientation and mobility of people with visual impairments, both outdoors and indoors, from different perspectives. In the following paragraphs, we present some of these research projects.

Part of carrying a trip to an unknown location is considering the use of transportation. Several studies such as Golledge et al. (1997), Marston et al. (1997), Montarzino et al. (2007), Penfold et al. (2008), Park and Chowdhury (2018) address the issue of public transportation for travelers with visual impairments. Our focus is less on the transportation and more on general orientation and mobility, i.e. finding the way, staying on the right path, travelling safely and knowing what's around.

Banovic et al. (2013) conducted interviews with specialists in orientation and mobility (O&M) and a qualitative study with nine people with visual impairments. The study included interviews and observations of in-situ navigation in two unfamiliar locations, both outdoors and indoors. The article focuses on the strategies that individuals who are visually impaired use to learn about their environments, the challenges they face and the underlying requirements associated with learning about an environment. They found that there are four types of environmental information that people with visual impairments need: high-level, safety, navigation information and points of interest. They also conclude that "users not only learn information to satisfy their immediate needs, but also to enable future opportunities - something existing technologies do not fully support" (Banovic et al. 2013).

McIntyre (2011) observed eight people with visual impairments while navigating through a complex building in order to draw conclusions about the wayfinding strategies of this user group. She finds that each journey depends on what the user tries to achieve (find an address or explore) and the user's own approach and learning style. Her conclusion is that the main cause of wayfinding problems is the lack of communication and not the complexity of a building. According to her, users need five types of communication during a journey: "1. Identification, 2. Orientation, 3. Navigation, 4. Warning and 5. Instruction" (McIntyre 2011). She also draws conclusions about related architectural design issues.

Gomez et al. (2016) conducted a literature survey, an interview with three and a survey with twenty six employees with visual impairments in Chile. Similarly to McIntyre (2011), their conclusion is that one common problem of this target group in terms of mobility and orientation is the *lack of information*. The participants additionally stated that Braille is not useful to them and that "it took them a long time to get to know the workplace [...]. Some of the participants even disclosed that they only knew how to get to the essential areas such as the workstation, kitchen, toilet, lifts, and the way out" (Gomez et al. 2016). This is in line with the findings of Branham and Kane (2015). Thus, people with visual impairments often know very little about the buildings they visit regularly.

In a study that conducted phone interviews with 30 people with visual impairments from the US it was found that most of them "relied on sighted guides when navigating unfamiliar indoor spaces" (Williams et al. 2013). In another study that interviewed 20 visually impaired people from the US and South Korea about travelling to unknown destinations, many of the participants "reported using private transportation services, sighted guides, or going along with friends or family. [Some] might call ahead for directions, [...] others seek information from other visually impaired friends or trainers" (Quinones et al. 2011). This shows the lack of information that people with visual impairments can access independently.

Information needs (accessibility features). When it comes to the information that people with visual impairments need when travelling, there is no consensus about the terminology used. Most projects use their own terminology to define parts of the information needed, depending on the scope of the research conducted. Parker et al. (2021) found in a review of 35 articles that "26 of the studies mentioned landmarks but not always as a direct term. For example, sometimes other terms were used such as 'points of interest', 'features', 'environmental features' or 'action points'" (Parker et al. 2021). Kammoun et al. (2010), however, differentiate between *landmarks* which they define as points used to confirm one's position on the route and *points of interest (POI)*, which are potential destinations, but also interesting points on the route that help to better understand the environment. We call all the information needed by people with visual impairments *accessibility features*, which include at least landmarks, POIs and obstacles in the environment. Thus, we consider objects or things that influence both positively and negatively the travel experience. We consider that it is difficult to draw a line between POIs, landmarks and obstacles, as one object can be one, two or all of them at once. A trash bin, for instance, can be a POI when the user wants to throw something away; it can be a landmark when the user knows that he must turn right after the trash bin; and it can be an obstacle that one bumps into.

Koutsoklenis and Papadopoulos (2014) conducted several user studies with 32 people with visual impairments - several users participated in more than one study - and came up with a list of 37 haptic cues used for outdoor wayfinding. Of these, the 10 most significant are: changes in texture of the walking surface, sidewalk, bus stop, slope, curb ramp, wall, parking post, traffic light, flower bed and pothole.

Bredmose et al. (2023) let people with visual impairments rate 34 physical elements from the urban space, such as: handrails on stairs, tactile and visual delineation of walkways, attention patterns on tactile walking surface indicators at bus stops or kerbs at refuge islands. All elements were rated important by either blind people or by people with low vision. For blind people, for instance, tactile elements were best rated, while for people with low vision, contrasting colors on elements were most often rated as important and very important. The element with the highest rating for all participants was “audible beacons in signal controlled crossings”, pointing to the importance of audio signals in safety critical situations such as crossings.

While outdoors there is more knowledge about the information that people with visual impairments need while travelling - due to the existing systems such as BlindSquare (MIP 2020), research projects such as InMoBS (Vollrath et al. 2015) or m4guide (Jaunich 2014, Projekt m4guide 2017) and research articles such as Koutsoklenis and Papadopoulos (2014), Bredmose et al. (2023), Norgate (2012) - indoors the situation is less clear.

Some research projects, such as Thapar et al. (2004), do mention indoor accessibility features. They measured “functional access [...] of 30 public buildings and facilities based on task performance” (Thapar et al. 2004) with one legally blind person and two people with mobility impairments. They found that for the person with visual impairment, signage, lighting, confusing building layout as well as amenity facilitators (such as lowered telephones, drinking fountains and soap and paper towel dispensers) are of great importance for the accessibility of a building. However, this was a pilot study and a quantitative evaluation with more users is necessary to confirm the results. Moreover, such projects only name few accessibility features and do not get near to defining an exhaustive list.

One study highlights the importance of indoor accessibility features (here called ‘landmarks’) for people with visual impairments: “participants repeatedly reported their urgent need to be provided with the locations of important landmarks around them when they enter unfamiliar buildings” (Alkhanifer 2015). We address this need and compile, in Section 3.1, a list of indoor accessibility features.

Electronic Travel Aids. We have seen that people with visual impairments, and especially blind people, rely mainly on others for travelling. They wish to travel more independently, and one solution is using tactile maps, while the other one is to use electronic travel aids.

According to Espinosa et al. (1998), tactile maps are an effective means for people with visual impairments (VI) to learn the spatial structure of an unknown urban area: “tactile maps can potentially increase independence and autonomy of people with VIs while also supporting O&M practice” (Espinosa et al. 1998). Others propose a tablet App “for providing on-the-go access to audio-tactile maps of unfamiliar indoor venues to individuals who are blind or visually impaired. [...] Results indicate that the audio-tactile display improves survey knowledge” (Adams et al. 2015). Goldschmidt (2018) also takes a look at tactile maps from the point of view of Orientation&Mobility training. She concludes that “tactile maps and models are good alternatives to visual preview but the ability to use them effectively and efficiently has to be developed” (Goldschmidt 2018). Moreover, tactile maps are hardly available, as the equipment to produce them is very expensive and requires most of the time trained sighted personnel who creates or adapts the tactile material.

Thus, the more practical alternative is to use electronic travel aids instead of tactile maps. A study showed that people with visual impairments would like to use technical applications while travelling (Gupta et al. 2020).

According to the research of Walle et al. (2022), electronic travel aids for people with visual impairments must contain three modules in order to be helpful to this target user group:

1. The **navigation or wayfinding** module, which must take into account the user’s location and orientation throughout the route. It must also include a description of the route, which must be chosen taking into account criteria such as safety and user preference. This module should also provide information about the texture of the walking surface throughout the route. It should work both indoors and outdoors independent of the light conditions. Among the non-functional requirements Walle et al. (2022) mention: real-time analysis, robustness and accuracy.
2. The second module should perform **object detection** of both static and dynamic objects, for two purposes. First, for obstacle avoidance, the module should convey to the user the nature and location of the objects, including distance. Secondly, object detection should be performed for scene description called upon user demand, to help the user understand the environment and build a cognitive map.
3. Finally, the third module is **the interface** - both input and output. It should be chosen based on algorithmic methods, user preference and the wearability of the input and output devices.

Currently, electronic travel aids are divided between outdoor and indoor systems. Comprehensive solutions do not yet exist. Outdoors there are some systems, whether mainstream, such as Google Maps (Google 2018), or specifically designed for this target group, such as BlindSquare (MIP 2020). However, even the ones specifically designed for people with visual impairments “often fail to address the real needs of the users” (Dramas et al. 2008).

Among all electronic travel aids, Apps that handle accessible Points of Interest (POI)-based navigation and accessible turn-by-turn navigation are the most relevant to us. *BlindSquare* (MIP 2020) and *Lazzus* (Nue 2020) for instance are accessible Global Navigation Satellite System (GNSS) Apps that convey to the user information about the current address (on demand), POIs around the current location and also street intersections. *Lazzus* additionally includes information about the location of pedestrian crossings, traffic lights, zebra crossings and stairs. *ViaOpta Nav* (Novartis 2015) is a free turn-by-turn navigation App developed specifically for people with visual impairments. It includes information about intersections as well as crossings, traffic signals and tactile pavings. *Seeing Eye GPS* (Sen 2020) and *SeeingAssistant Move* (Transition Technologies S.A. 2018) combine both POI and current location information with navigation instructions in an accessible way. *Seeing Eye GPS* additionally provides descriptions of intersections, but does not support sidewalks (roadside). Another App specifically designed for people with visual impairments is iMove (Kacorri et al. 2016, 2017), which provides information about the current location, POIs and geo-notes, which are notes created by the user and associated with a particular geographic location. But, as mentioned before, none of these apps manages to impose itself and address the needs of the majority among people with visual impairments.

Indoors, the situation is even more challenging. Although there are several navigation systems for indoors, holistic commercial solutions are basically inexistent. Some of the systems proposed (Murata et al. 2018, Kacorri et al. 2018) use Bluetooth beacons that must be installed in the buildings and maintained. Others (Kim et al. 2016) use only technologies available in the smartphone, like Bluetooth and a compass. Yet others employ computer vision techniques to detect: markers in a hospital (Bashiri et al. 2018), lists of objects (Mekhalfi et al. 2015, Mekhalfi et al. 2017), indoor navigational signs (Kunene et al. 2016). Fallah et al. (2013) give an overview of existing systems that use RFID, IR, Ultrasound, Bluetooth beacons or barcodes. They conclude that systems did not achieve large-scale deployment due to cost, accuracy and usability, including *the lack of a robust feedback modality*. This inspired us to address the needs of the users and propose feedback modalities based on speech and sound in Chapters 4 and 5.

2.2. Output Interfaces for People with Visual Impairments

A recent review from 2022 of 33 projects featuring artificial intelligence and vision-based solutions for people with visual impairments found that “the most popular choice is audio-feedback through speech instructions” (Walle et al. 2022) - in 70% of the studies. Another 13% of the papers featured non-speech sonification methods, while 17% used haptics as feedback modality[1], most often in the form of vibrotactile feedback. The haptic output devices mentioned in the sources analyzed by Walle et al. (2022) are: vibrating motors integrated into a white cane, vibrating belt, the vibrations of a smartwatch or of a smartphone and Braille interfaces. Other journal articles that review navigation solutions for people with or without visual impairments, such as Plikynas et al. (2020), Kandalan and Namuduri (2020), Fallah et al. (2013), Huang and Gartner (2009), also found that most output modalities used, besides visual, are auditory and vibrotactile.

This summarizes well the current trend in *mobile* user output interfaces for people with visual impairments: speech, non-speech sonification and vibrotactile interfaces are currently by far the most feasible choices. Other haptic interfaces exist, such as: pin-matrix displays (Prescher et al. 2018), force-feedback devices (Ritterbusch et al. 2012), haptic gloves (Zimmerman et al. 1986) or exoskeletons (Yang et al. 2008), but these are not mobile, so they cannot be used as interfaces during travel.

The advantages of haptic interfaces over audio ones include the fact that haptic interfaces do not mask the environment, are language-independent and can be faster than speech. But mainstream haptic interfaces existent nowadays are unfortunately either very expensive - such as Braille displays, vibrating

---

[1] While the definition for ‘haptics’ seems to be controversial or at least inconsistently used when it comes to interfaces for people with visual impairments, we use in this thesis the definition of the term as given in neuroscience: “haptics refers to the sensory inputs arising from receptors in skin, muscles, tendons, and joints that are used to derive information about the properties of objects as they are manipulated”. Haptics includes thus both the tactile (touch) and proprioceptive (kinesthetic) sensory modalities (Jones 2009, Pẽrez and Santĩs Chaves 2016). For instance, experiencing the texture of an object held in the hands relates to the tactile modality, while assessing its weight relates to kinesthesia.

belts -, bulky or cannot convey as much information as auditory ones, especially when compared to speech. This is the case for instance when using the vibration of a conventional smartphone, which usually has just one vibrating motor inside.

The advantages of audio interfaces, on the other hand, include the fact that they can convey a lot of information, no matter how complex; additionally, the output devices are cheap and finally, audio interfaces can be easily programmed on almost any platform, especially mobile ones such as smartphones. Thus, while haptic interfaces should be considered at least complementary to audio interfaces, in this thesis we address audio interfaces alone.

2.2.1. Speech Auditory Displays

Research in speech synthesizers has been conducted since the 1930s (Klatt 1987). Text-to-speech is currently widely used by the masses, for instance in car navigation systems, such as those based on GPS. For people with visual impairments, and especially blind people, speech synthesis opened the doors to technology use, from PCs to modern smartphones. Screen readers, pedestrian navigation systems and even products of daily living such as talking weighing scales make the life of people with visual impairments easier. But while the use of technology at home is somewhat safe, its use on the street is not without dangers. The speech output of mobile assistive systems must be very carefully designed, so as not to interfere with the environmental sounds - which are used by people with visual impairments to travel safely. At the same time the amount of information that needs to be conveyed through speech synthesis is very high. Thus, a balance must be found between conveying too much and not enough information. But, while the development focus has always been on speech synthesis and providing speech that sounds as clear and natural as possible - the structure and content of the messages has been less considered.

When designing auditory displays, one must consider, among others, the potential, but also the limitations of the auditory memory. “The auditory system contains a brief auditory store (‘immediate’ or ‘echoic’ memory) where a crude representation of the sensory stimulus is maintained, normally for no longer than 2 seconds. This store makes it possible to retain a sound temporarily in order to make comparisons with later-arriving sounds, as well as to process simultaneous stimuli in a serial fashion. When stimuli are processed in more detail (such as the semantic processing of speech, or the learning of a sound pattern in the environment), there is the possibility for more permanent, categorical representations and long-term storage” (Peres et al. 2008). In this case, learning plays an important role, especially in the case of non-speech sonification, but also in the case of speech, as vocabularies can be very different even between people in the same country or even city. There are many factors affecting the understanding of speech instructions, including: age, education, local dialect, migration background or hearing impairments. This is another reason why the speech output of electronic travel aids must be carefully designed in order to address such a diverse population, reduce learning and foster early adoption of technology.

The speech feedback of mainstream navigation apps was initially designed for use in cars, or, at best, for sighted pedestrians. Thus, it does not include all the information that satisfy the navigation and wayfinding needs of people with visual impairments (Banovic et al. 2013) such as: the presence of sidewalks, roadside, details about pedestrian crossings or changes in ground texture. Instructions are given when to turn, but getting on the correct side of the street or crossing an intersecting road is left to the pedestrian. Moreover, users must familiarize themselves with each new system including its output modalities. These are often very different, verbose, not customizable or hardly adapted to the needs of people with visual impairments (Allen 2000).

Among the research projects that focus on the speech instructions for people with visual impairments during travel, we count the ones of Nicolau et al. (2009a,b). They analyzed how people with visual impairments verbalize routes and used this information to define a grammar used for indoor purposes that fulfills the needs of the users.

Some projects (Jacquet et al. 2004, Ivanov 2017) propose modelling the indoor environment for navigation and orientation purposes. Their models include "geometric, topological and semantic information [such as room owner or access restrictions] about the building elements and objects" (Ivanov 2017) and name important landmarks, obstacles and hazards relevant to the target group. Goyal et al. (2019) and Madugalla et al. (2020) generate textual descriptions from floor plans.

In the Mobility project (Spindler et al. 2012), the main goal was to provide descriptions of indoor environments to support people with visual impairments in both wayfinding and exploration. Additionally, descriptions of the route, i.e. of how to walk through the current area to reach a specific target point are provided. A pilot study with 6 blind users in a large airport yielded good results while at the same time uncovered many possible improvements to the contents and timing of the speech instructions.

Van der Bie et al. (2019) propose Sidewalk, an *English* wayfinding syntax for people with visual impairments. Sidewalk has a consistent structure and provides detailed wayfinding instructions, short instructions and alerts.

In Nicholson (2010) a framework is proposed that describes how to build collaborative websites in which users with visual impairments write, edit and share route descriptions. Based on these user descriptions, which are written in natural language and are thus unstructured, the author proposes a method to extract information such as landmarks and their position. Finally, using graph algorithms, completely new and structured routes are generated from the unstructured route descriptions given by users, complementing thus the route collection.

In Kamikubo et al. (2020), a comparative study is conducted to assess the performance of trained versus untrained remote sighted guides. Based on the results, guidelines for verbal description methods are formulated. While they are meant to be used in the context of remote sighted guidance, some of them could apply to speech-based interfaces as well.

2.2.2. Non-Speech Auditory Displays

Sonification is not a new concept, but, despite its age, it is a hardly explored topic. This is due, on the one hand, to the fact that vision dominates over audition in adults (Hirst et al. 2018), which explains why visual displays prevail over auditory ones. On the other hand, sonification is a highly interdisciplinary field (Hermann et al. 2011), making it difficult for researchers to get all the expertise and funding needed. Yet, in few areas, sonification has found broad acceptance. Applications for sighted people include quick alerts in smartphone apps, for instance when a message arrives, or in home appliances such as refrigerators and ovens. Other well known sonifications include the parking assistance in cars which beeps faster the closer the car gets to an object. However, sonification can be used for much more than just alerts. According to Hermann et al. (2011), auditory displays fulfill 4 main functions: (1) alarms, alerts and warnings; (2) status, process and monitoring; (3) data exploration; and (4) art, entertainment, sports and exercise. The Sonification Handbook (Hermann et al. 2011) also names 5 sonification techniques, which will be explained shortly:

- *Audification* is “the direct translation of a data waveform into sound. This often requires that the data wave be frequency-shifted into the audible range for humans, or time-shifted (slowed down or sped up) to allow for appropriate inspection by the listener” (Walker and Kramer 2004). Examples of data that are well suitable for audification are EEGs (electroencephalograms) or seismic data.
- *Auditory icons* are “the auditory equivalent of the visual icons” (Hermann et al. 2011). Thus, they are intuitive sounds that map to an object or an action. An example of auditory icon is a bicycle’s bell to represent a bicycle.
- *Earcons* are short, structured sounds that map to some given data. Earcons are somewhat similar to auditory icons, except that there is no initial relationship between the sound and the data mapped. This means that the earcons must be first learned, thus, unlike auditory icons, they are not immediately intuitive. An example would be the ringing tone of the smartphone, but earcons have been used long before smartphones or even computers existed. They were used for instance in the 19th century in wartime to broadcast to the troops information about incoming mail, mealtime or enemy attack. For this purpose, bugles (a kind of small trumpet) were used. The arbitrary nature of the mapping made the communication safe from enemies (Hermann et al. 2011).
- *Parameter mapping sonification* (PMSon) means, as the name suggests, mapping parameters of the data to parameters of the sound. “Since sound is inherently multidimensional, PMSon is – at least in principle – particularly well suited for displaying multivariate data” (Hermann et al. 2011). An example is the parking alert system that beeps faster the closer the car is to an object. Thus, the distance to an obstacle is mapped to the sound’s tempo. Additionally, objects on the left side of the car are played on the left speakers and objects on the right on the right speakers. Thus, object position on the x-axis maps to sound panorama (panning).
- *Model-based sonification* “is defined as the general term for all concrete sonification techniques that make use of dynamic models which mathematically describe the evolution of a system in time, parameterize and configure them during initialization with the available data and offer interaction/excitation modes to the user as the interface to actively query sonic responses which depend systematically upon the temporal evolution model” (Hermann et al. 2011). In short, the sonification model creates sound in response to user interaction. Model-based sonification “provides a means of experiencing data by interaction with data-driven objects. The linkage from data to sound is fully determined by the model and more complex than the mapping used

in [parameter mapping]" (Hermann 2002). Thomas Hermann is the one who introduced the topic of model-based sonification and refined it in his dissertation (Hermann 2002). Here, he also gives examples of model-based sonifications.

Among the sonification techniques described above, parameterized auditory icons - meaning, a combination of auditory icons and parameter mapping - are most suitable for sonifying objects in the environment and their position, due to their intuitiveness. Thus, they are most interesting to us and are the sonification types addressed in this thesis. In the following paragraphs we give a little more insight into auditory icons.

Auditory icons were introduced for the first time by William Gaver in 1986 in computer interfaces, as supplements to graphical user interfaces (GUI). Gaver called auditory icons "caricatures to naturally-occurring sounds" (Gaver 1986). An example of auditory icon in the context of computer interfaces is the paper crumpling sound for the recycle bin. "An 'icon' does not imply a literal, pictorial or recorded mapping, instead that its characteristics are causally related to the things it represents. So an icon may be a sketch, outline, or caricature; an auditory icon may be a recorded sound, or one synthesized to capture important features of an everyday sound. What is important is that the attributes of the representation convey information by virtue of their causal relations to the attributes they represent" (Gaver 1989). Gaver categorized auditory icons based on the kind of mapping between the object or action and its associated auditory icon. Thus, according to Gaver (1986), auditory icons can have three types of mappings:

- *Symbolic* mappings are "essentially arbitrary, relying on social convention for their meaning. Telephone bells, sirens and stop signs are examples of symbols" (Gaver 1986). Thus, they are intuitive for people who know them already, namely who encountered them in their everyday life, but for the rest, they must be learned. Thus, they are culture-dependant.
- *Metaphorical* mappings "make use of similarities between the thing to be represented and the representing system: they are not wholly arbitrary, yet they do not depend on physical causation" (Gaver 1986). An example would be using several notes in rising pitch for representing height.
- *Nomic* or iconic (Gaver 1989) mappings are most intuitive, as "their meaning depends on the physics of the situation [...]. The representations are images of the information." (Gaver 1989). In other words, nomic mappings are those between a sound and its source. Gaver (1986) gives as example for nomic mapping the mailbox sound for arriving mail.

According to Gaver, "nomic mappings should be relatively simple to learn, metaphorical mappings somewhat harder, and symbolic mappings the most difficult" (Gaver 1986). However, symbolic mappings can be assumed to be already learned in a given culture by the adult population. Thus, depending on the target end users, this kind of mapping could also be intuitive. Finally, Gaver (1986) also mentions that a mapping can fall between the three categories, or move between them depending on the user's background, context of use and learning of the auditory icons.

Sonification for people with visual impairments in the travel context. For people with visual impairments, sonification is used by many applications, for instance to make graphs accessible (Walker and Mauney 2010, Vines et al. 2019) or to explore graphics or virtual maps (Banf and Blanz 2012, Schauerte et al. 2015, Geronazzo et al. 2016). In the travel context, sonification is used to indicate rotation instructions in indoor navigation environments (Ahmetovic et al. 2019), to guide the user or to support navigation tasks (Parseihian et al. 2016, Wilson et al. 2007, Katz et al. 2012, Ziemer and Schultheis 2018, Vasilijevic et al. 2018, Urbanietz et al. 2019, Aziz et al. 2019) and to present nearby features, points of interest and fixed obstacles (Wilson et al. 2007).

Ahmetovic et al. (2019), for instance, investigated intermittent sound, amplitude modulation and musical scale to give the users hints about the direction in which to rotate their bodies. In a later work, Ahmetovic et al. (2023) investigated the use of sonification for conveying ‘turn’ and ‘walk straight’ instructions.

Wilson et al. (2007) proposes the prototype of a system for wearable audio navigation (SWAN). Although the work is dated 15 years ago, it is probably one of the most comprehensive and suggestive in the area of sonification interfaces for people with visual impairments in the travel context. The interface is composed of non-speech sounds only. The hardware is composed of a portable computer, audio processor and a position and orientation tracking device. The entire hardware fits into a shoulder bag or small backpack. The sound is transmitted through bone conducting headphones - this research project is probably the first one to introduce them. SWAN computes a route to the destination composed of segments and nodes, called waypoints. A spatialized beacon sound continually guides the user to the next waypoint. When approaching it, the tempo gets faster, similar to the car parking assistant. When the waypoint is reached (meaning any point on a 0.75 meter radius around it), a “subtle success chime” (Wilson et al. 2007) is played. If the user overpasses the waypoint, the timbre of the beacon changes. Besides navigation instructions, SWAN concomitantly conveys information about fixed objects in the environment, which are stored in a GIS database. For that, a combination of auditory icons, earcons and spearcons is used. Additionally, the sound is spatialized, so the position of the objects is also conveyed. Unfortunately, no further information about the object sonification is provided. It also seems like the system has never hit the consumer market. While the reasons are unknown to us, we can assume that the GPS inaccuracy in cities and the still rather bulky hardware might have been the cause.

Object Sonification. In Yang et al. (2019) a sonification method for conveying objects in the environment is presented. The system consist of 11 cameras installed at fixed indoor locations that track both the user and the objects to be sonified. For that, the user must wear a helmet. A laptop is used for the computations and the sound is sent to the user’s Apple earpods wirelessly. One object is sonified at a time, so one sound is used for all objects. An average head-related transfer function (HRTF) is used for conveying 3D sound to the user. The system does not scale well due to the installed cameras, so it cannot be used in the travel context in general, but it could be used in given indoor spaces. Moreover, it was intended for sighted people, so it will have to be investigated whether it is also appropriate for people with visual impairments.

Another project also investigated 3D sound, this time in combination with target localization - which can also be seen as an ‘object’. They said that they were focusing on “3D audio rendering methods that combine auditory target localization with natural perception of the environment” (Dramas et al.

2008). The authors intended to design a device with cameras mounted on glasses that identify objects and their location with respect to the user. At the same time, they intended to convey the location of these objects through virtual sounds. Whether they managed to do so, it is unclear, as no results are described in the mentioned paper. The problem of 3D sound for blind people is discussed in more detail in Section 5.3.

Conclusion. Literature shows that people with visual impairments rely most often on other people when travelling, but that they wish to be more independent and obtain information autonomously. To achieve this goal of independent information retrieval, mobile assistive devices, called electronic travel aids, could be the solution. But for them to be effective, one needs more, namely:

1. Freely available information - what information precisely, we address in Section 3.1, as well as in the evaluations performed throughout Chapters 4 and 5, in which we often ask users about the information they need;
2. Mobile solutions that access this information;
3. An interface for the mobile solutions to convey the information. *How* to convey it in an appropriate and effective way for this user group, we cover in Section 3.2, where we specify the non-functional requirements of the interface. The interfaces themselves are the main topics of Chapters 4 and 5.

Our focus is on points (1) and (3), namely the information needs and the interface. While point (2) is not the focus of this thesis, we do implement our interfaces in existent prototypes as proof of concept in Subsections 4.1.2, 4.2.4 and Section 5.1.

3. Requirements for Accessible Travel

In this chapter, we address the question of users' information needs and non-functional interface requirements for mobile travel assistance systems. Or, in other words: *what* information do people with visual impairments need when travelling and *how* should this information be conveyed?

Requirements are very important for understanding how an interface should be built. According to Shneiderman et al. (2018, p. 132), "soliciting, capturing, and specifying user requirements are major keys to success in any development activity." On the other hand, not all requirements can be determined from the start, so often, systems must be first implemented and evaluated in order to find out how to improve their usability (Dix et al. 2004, p. 235). While functional requirements specify the behavior of a system, or *what* it is supposed to do, non-functional requirements define *how* the system should be, or the way in which functionality should be provided (Dix et al. 2004, p. 229), but "without being tied to a specific action or behavior" (Shneiderman et al. 2018, p. 132).

Before establishing the requirements of a system, one must first identify the users' needs. Requirements are ways of addressing users' needs taking into account constraints such as, for instance, technology affordances (i.e. what can be achieved with technology at a given point in time). One major type of user need relates to information: what information need does the system address? In our case, what information do people with visual impairments need when travelling? If this question has been addressed, up to some extent, for outdoor environments, thanks to a more advanced state of the art of technologies that assist outdoors - for indoors, a comprehensive list of information needed is missing.

In section 3.1 we first identify the information needs of users when travelling into unknown buildings and create an initial list of what we call *indoor accessibility features*. In section 3.2 we then define an initial set of general, non-functional requirements for auditory displays of mobile assistive systems for people with visual impairments.

3.1. Indoor Accessibility Features

This Section is based on our ICCHP publication Constantinescu et al. (2022a) © Springer.

In this section, we assess users' informational needs with respect to indoor travel scenarios, or, in other words, we answer the question: what information is important to people with visual impairments when travelling through unknown buildings?

In order to find out how people with visual impairments travel indoors and what their needs are, we conducted an online survey[1] with 106 people with visual impairments with ages between 8 and 77 years (two participants were underage) and a balanced male-female distribution. The majority of the respondents came from Germany, 13 from other European countries and only 6 from the rest of the world. Of the participants with visual impairments, 40% had low vision (visual acuity less than 30%) and 60% were blind (visual acuity less than 2%).

According to this survey, when travelling indoors, both people with blindness and low vision find their way mainly by asking other people in the building for directions. Almost a third of the blind participants said that they travel accompanied by an assistant. However, both user groups (with blindness and low vision) would prefer to travel more independently inside of buildings, for instance by making use of textual descriptions of the building or a digital map. For that, assistive applications for indoors would be helpful. But to date, there are hardly any such applications available, as: (1) localization in buildings for the special needs of people with visual impairments requires an infrastructure which is difficult to setup and/or to maintain; (2) there are hardly any indoor maps publicly available; and (3) there is no widely accepted, exhaustive understanding of the information needed by people with disabilities inside buildings. Through a literature review, we want to close the last gap, related to the information needs, by creating a list of indoor accessibility features. By *accessibility features* we mean objects or things that help or impede people with disabilities during travel. Examples range from structural elements such as form of the building, type of entrance door, to objects and their location such as stairs, doors and elevators, and finally to features such as quality of lighting or smells at a certain location and during a given time of the day.

3.1.1. Literature Review on Accessibility Features

We performed a literature review in order to find the accessibility features that are relevant for people with visual impairments when travelling indoors. Initially, we also wanted to compare the results with the accessibility features needed by people with *mobility impairments*. However, since this is out of the scope of this thesis, we focus here on visual impairment only, but do mention mobility impairment when it is needed for reproducibility and completeness[2].

A first and quick review of literature conducted by Vanessa Petreausch and Angela Constantinescu in 2020 yielded 15 studies containing indoor accessibility features for people with visual impairments or with impairments in general (including visual), see 1st row in Table 3.2: Charitakis and Papadopoulos (2017), Saha et al. (2019), Jeamwatthanachai et al. (2019), Gomez et al. (2016), Jaunich (2014), TUD (2011), Thapar et al. (2004), Park and Chowdhury (2018), Holfeld (2011), KIT and IOSB (2015), Grohmann et al. (2018), New England ADA Center (2016), Pérez et al. (2017), Alkhanifer (2015), Stude et al. (2012). Since we were not confident that we had found all relevant features, we decided to

---

[1] The questionnaire was mainly implemented by project partners from TU Dresden, especially Christin Engel. Angela Constantinescu was involved in the process, but to a much smaller degree than the TUD partners.

[2] As a side note, we found very few studies dealing with accessibility features in the context of mobility impairment with the searches performed. From the literature found, we noticed that the word “mobility” is often used in the context of *visual impairment* to mean restricted travel behavior which is due to the visual impairment.

carry out a second literature review, this time a systematic one after Kitchenham and Charters (2007). We started by searching in Science Direct, IEEExplore, SpringerLink, ACM DL, Google Scholar and Google Search to see whether a systematic literature review had already been performed. The keywords searched were: systematic, review, orientation, accessibility, indoor. Given that no relevant results were found, we concluded that such a systematic review most likely does not exist. We thus conducted a systematic literature review ourselves in April 2020.

Keyword analysis. We first performed an analysis of already known literature that is highly relevant to our topic, including projects previously conducted at our institute (Project ATMaps, **?**). The purpose was to come up with a list of keywords to be used in the search for literature. We first analyzed 6 studies - books, papers, project reports (TUD 2011, Project ATMaps, Alkhanifer 2015, Pérez et al. 2017, New England ADA Center 2016, Stude et al. 2012). We extracted keywords from the title and abstract or problem definition, project description or introductory section. In two cases, we translated keywords from German to English. We identified keywords by removing linking and other irrelevant words from the text, i.e. most pronouns, prepositions, conjunctions, articles, interjections and adverbs; many verbs such as "to be", "to have", "to prove"; and few general nouns and adjectives which were not related to our topics of interest, such as "individual", "method", "design" and "user". We then merged keywords with the same word root, such as "access", "accessible" and "accessibility", and counted the number of sources in which each compound keyword appeared. Finally, we excluded keywords that: (1) limit the scope of the search, such as the word "public", because we also address private buildings; (2) keywords that were synonyms with other, more prominent keywords, such as the word "assistive", which only appears in one study and is synonym with "accessibility", which appears in three studies. The keywords thus obtained were: orientation, access*, barrier, mobility, blind, building, indoor. Since synonyms are still present, we grouped them together with the logical operator OR. Thus, the keywords search can be summarized as follows:

orientation AND (access$*$ OR barrier OR blind OR mobility) and (building OR indoor)

where *access*$*$ stands for access, accessible, accessibility.

Searching with these keywords yielded poor results, namely very few studies whose title and abstract seemed relevant. Thus, we performed a second keywords analysis, including more relevant literature. This time, we analyzed 7 studies (Gomez et al. 2016, Sato et al. 2019b, Goldschmidt 2018, Charitakis and Papadopoulos 2017, Alkhanifer and Ludi 2014, Pérez et al. 2017, New England ADA Center 2016), 2 of whom were also used in the previous keywords analysis and 5 new, found either in references of the previously known literature, as a result of the first keywords search or with alternative searches. These studies had a higher relevance to our topic than the previous ones. This time, we analyzed relevant parts of the body text of the studies, besides title and abstract, to retrieve relevant keywords. For instance, when the study was more generic, and only one part of it was relevant, then only the text in this part was analyzed instead of title and abstract. We gathered all the keywords that seemed relevant to the topic. We also included some keywords that seemed very relevant to the paper or where we were not sure. We removed any linking words and general terms (such as "existing", "design", "improve"), similar to the previous keywords analysis. When keywords appeared more than once in a study, they were counted twice (were given a weight of 2) - otherwise, they were counted once (were given a weight of 1). Similar terms were grouped together in clusters. Words within the same cluster were connected with OR, and clusters were connected with AND. This is the resulting long search:

(navigation OR orientation OR mobility OR wayfinding OR spatial) AND
("visual impairment" OR "low vision" OR blind OR disability OR wheelchair OR accessible OR assist OR "mobility impairment" OR "physical impairment" OR "physical disability") AND
(indoor OR building OR facility OR workplace OR office) AND
(feature OR landmark OR "situation awareness" OR semantic OR taxonomy OR information OR clue OR requirements OR factors) AND
unfamiliar AND
map AND
safe.

We also created a shorter version of the search, by including from each cluster only the keyword or keyphrase that appeared most often (had a weight of at least 4) in the studies analyzed, and also removed the cluster that contained only one keyword with a weight of 2. The resulting *short search* gave fewer irrelevant results than the long search, but similar relevant ones, both in number and content. We thus decided to keep the short version only, which is also our **final keywords search**:

orientation AND ("visual impairment" OR "mobility impairment") AND indoor AND requirements AND unfamiliar AND map

Databases searched. According to Bhowmick and Hazarika (2017), four of the most prominent databases in the field of "Assistive Technology for the Visually Impaired and Blind People" are: *Elsevier ScienceDirect*, *IEEE Xplore*, *ACM Digital Library*, and *Thomson Reuters Web of Science*. We added *SpringerLink* to the list, as the proceedings of ICCHP - an important conference in the field, at least for the European area - are published in Springer. Searching in these 5 databases using the short keywords search, we found a total of 397 studies to be further analyzed.

Analysis of title, abstract and full text. In the next step, titles and abstracts of these studies were analyzed. Studies that obviously did not address orientation and/or mobility, navigation, or spatial awareness for people with visual impairments or mobility impairments[3] in indoor environments were discarded. Moreover, studies for which we could not get access to the full text were also excluded at this stage. Applying these criteria, 45 studies were found to be possibly relevant: 32 from ACM DL, 12 from SpringerLink and 1 from ScienceDirect.

---

[3] The initial search also addressed people with mobility impairments. However, these were underrepresented throughout the entire study, and it is beyond the scope of this thesis to report them here. In the second keywords analysis, for instance, all studies addressed people with visual impairments and only two studies addressed people with visual *and* mobility impairments. In the keywords obtained after the second keywords analysis, only the word "wheelchair" was related to mobility impairment, and had the lowest weight of 1. Few relevant studies related to mobility impairments were found in the end, namely only 4.

(a) Number of studies per database

Database	Number of studies
ACM DL	238
SpringerLink	128 (20 preview-only)
IEEExplore	0
ScienceDirect	31
Web of Science	0
Total	397 studies

(b) Number of studies per group

Search	ALL	VI
Simple	4	9
Systematic	-	7
Systematic-alt.	-	2
Total	4	18

Table 3.1.: **(a)** Number of studies found in the five databases. **(b)** Number of relevant studies found for each target group in the three searches performed: simple, systematic and systematic review with alternative keywords. ALL: all disabilities; VI: visual impairment.

The full text of these 45 studies was further analyzed according to the following inclusion/exclusion criteria:

Inclusion criteria: (1) Deals with visual or mobility impairment; AND (2) Mentions indoor features that affect orientation and/or mobility of the target group.

Exclusion criteria: (1) Does not mention any accessibility features; (2) Duplicate; (3) A more relevant study by the same authors or project was already included.

Results of literature review. After the full text analysis, only 6 studies Goldschmidt (2018), Banovic et al. (2013), Sato et al. (2017), Yang et al. (2011), Fallah et al. (2012), Williams et al. (2016) were found to contain accessibility features for people with visual impairments. Another study, McIntyre (2011), was found in the references during the full text analysis, and two studies, Ivanov (2017) and Alkhanifer and Ludi (2014), were found during the systematic literature review, but with alternative searches, i.e. using different keywords. We also added the 13 from the 15 sources found previously with the non-systematic search. We removed two studies, Stude et al. (2012) and Alkhanifer (2015), that were similar to other studies already included, or more appropriate studies from the same author had been included instead. In the end, the final list contained a total of 22 studies - see Table 3.2. This list contains architectural standards (Holfeld 2011), checklists (New England ADA Center 2016, Grohmann et al. 2018) and research projects. When reviewing the latter, we considered end user involvement. The papers chosen report comprehensive user studies with up to 53 (Sato et al. 2017), 95 (Alkhanifer and Ludi 2014) or even 120 (Charitakis and Papadopoulos 2017) people with visual impairments.

Since some of the most comprehensive studies we found addressed all disabilities (including e.g. mobility, hearing or cognitive impairments), without mentioning which features addressed which disability, we decided to create a single "complete" list for all disability types in order not to miss any features. In addition, starting from these results, we created a second list only for visual impairment - see Subsection 3.1.2.

Indoor accessibility features for all disability types. The complete list contains more than 820 indoor accessibility features and sub-features relevant to people with disabilities such as *door signs for next accessible entrance, automatic opening/closing mechanism of doors* or *car control buttons in Braille in elevators*. The features from German sources were translated to English, synonyms

Study	How was it found?
Charitakis and Papadopoulos (2017) Saha et al. (2019) Jeamwatthanachai et al. (2019) Gomez et al. (2016) Jaunich (2014) TUD (2011) Thapar et al. (2004) Park and Chowdhury (2018) Holfeld (2011) KIT and IOSB (2015) Grohmann et al. (2018) New England ADA Center (2016) Pérez et al. (2017)	Initial quick review of literature, including known projects. Two studies (Stude et al. 2012, Alkhanifer 2015) were removed after the systematic review, as similar studies or more relevant studies from the same author were included instead
Goldschmidt (2018) Banovic et al. (2013) Sato et al. (2017) Yang et al. (2011) Fallah et al. (2012) Williams et al. (2016)	Systematic literature review
McIntyre (2011)	Systematic review, in the references of one of the 45 papers whose full text was analyzed
Ivanov (2017) Alkhanifer and Ludi (2014)	Systematic review BUT alternative keywords search
Total: 22 studies	

Table 3.2.: The list of 22 final studies that contain indoor accessibility features for people with visual impairments, found with the simple and the systematic literature review

were merged and the resulting features were categorized under the following nine categories: *way to building, general building information, building geometry, facility daily needs, general help for orientation, change of elevation, change in ground height, movables* and *security* - see Table 3.3. We found that other ways of categorization are difficult. For instance, a categorization in *dangerous* and *useful* is not possible in our case, because what is useful to a person in a wheelchair might be dangerous to a blind person. Besides, assessments such as *accessible, dangerous, useful* are very individual, even within the same user category.

We noticed that some features were hierarchical (e.g. stairs – handrail – location of handrail), others were categorized according to the particular interest of their authors, so we decided to create a hierarchical representation: category-feature-property. The categories emerged naturally after looking at the features. We took from all sources those categories that matched our data and added new categories when needed. After structuring all data, further levels in the hierarchy appeared: sub-properties and even sub-sub-properties.

The building elements that have most features are *toilets* (97 features), *doorways* (90) and *rooms* (90), followed by *elevator* (64), *movables and furniture* (54), *signage* (53), *stairs* (45) and *floor / pathway*

Categories	Features	Few sample properties
General building information (28)	Formal information (24)	Address; Building number; Opening hours; Levels; Access restrictions
	Interior colors (1)	Conspicuous colors
	Light (3)	Visibility; Switch location
General help for orientation / Technical assistance (34)	Reference points (5)	Prominent place as starting point; Haptic; Olfactory
	Guidance system (2)	
	Signage (6)	Type; Language
	Plan (21)	Type; Location
Way to building (45)	Accessible Parking (12)	Number; Location; Dimension
	Exterior accessible route (13)	Condition: stable, firm, slip-resistant, cobbles; Dimension; Running slope; Cross slope
	Ramp (1)	
	Curb ramp / curb cuts (5)	Width; Level landing dimensions
	Lift outside (1)	
	Stairs (1)	
	Ground in front of entrance (4)	Condition: cobbles, asphalt; Tactile pavement to the entrance
	Description (1)	
	Way to public transport (6)	Distance to entrance; Name of station; Surface material
	Construction or work in the area (1)	
Change in Ground Height (30)	Step/s, handling of steps (1)	
	Elevated floor of teacher's desk (1)	
	Slope (2)	Running slope; Cross slope
	Ramp (26)	Where to; Slope; Dimension; Handrail: side, height
Building characteristics (273)	Doorways (90)	Name; Where to: building, toilet, passage door; Access: opening hours, key holder; Door lock; Carpet: thickness
	Floor (pathway) (40)	Access: public, restricted, open hours; Condition: slip-resistant, firm; Material; Dimension
	Wall (4)	Material: solid, glass, railings
	Obstacles (8)	Constructions; Movable; Location: hanging, overhead, ground; Colors
	Dangerous area / point (1)	
	Signs (5)	Braille; Contrast; Raised text
	Rooms / venues / offices (90)	Name; Address; Level; Purpose; Maneuvering clearance; Sockets; Installation; Balcony
	Windows (6)	Type: one-sided; How they open: automatic, heavy, tiltable; Shutters or blinds
	Technology (28)	Service number technician; Washbasin; Whiteboard: interactive; Over-head projector; Wi-Fi; Controls: height, ease of operation
	Stores (1)	Type: supermarket
Movables (55)	Furniture (7)	Type: hanging, seating, standing; Size; Material: glass
	Tables (6)	Type: fixed, folding; Dimensions
	Seating (12)	Number of seats; Seat rows number; Seating depth; Fixed seating
	Wheelchair space / seating (16)	Number; Clear line of sight; Route to accessible seating: width; Knee space dimensions: width, height
	Benches (7)	Maneuvering clearance at the end of the bench; Back support
	Blackboards (1)	Reflecting
	Food Service Lines (5)	Self-service shelf or dispensing service: height, shallow obstruction; maneuvering clearance
Change Elevation (129)	Elevator (64)	Number; Type: full size, LULA, passenger, freight; Number of doors; Access: public, keys; Opening hours; Door type: sliding, swinging; Floor number sign; Levels connected
	Escalators (10)	Direction of travel; Lanes; Handrail location
	Platform lift (9)	Lift controls height; Type of door: end, side
	Stairs / stairways (45)	Name; Type / shape: linear, u-shaped, open; Levels connected
	Moving walkway (1)	
Facility Daily Needs (153)	Services (21)	Opening hours; Type: information, service point, cashier, queue management ticket machine
	Toilet / toilet compartment (97)	Room id; Access: restricted, public, EURO key, gender, wheelchair; Lavatories: height, maneuvering clearance; Soap dispensers height, location; Flush control ease of use; Emergency call button
	Relaxation room (14)	Diaper change; Mirror; First aid kit
	Drinking Fountains (9)	Spout outlet height; Usage position: standing, sitting
	Public telephones (12)	Volume control; TTY height
Security (8)	Fire protection (2)	Fire extinguisher; Fire hose
	Fire Alarm Systems (2)	Flashing light signals; Audible signals
	Escape route (3)	Signposting without gaps; Illumination; Two-sense principle
	Security zone (1)	

Table 3.3.: Indoor accessibility features for all disability types. Once can see the structure (categories - features - properties). The categories and features are complete, while the "properties" column only contains a few sample items. In parentheses is the number of pairs "feature-property" contained in the category or number of properties contained in the feature.

(40). This might point out to the importance of these features. Around 16 % of the features describe dimensions such as height, width or length of objects, rooms and paths, which are particularly relevant for people using wheelchairs, but also partly for people with blindness. Knowing the height of a doorbell or a towel dispenser in a restroom, for instance, makes it easier to find its exact location.

Besides the obvious architectural features such as dimensions and physical properties of building elements and movables, we also found some other interesting accessibility features, such as *audio* and *olfactory cues*, *quality of lighting* or *high contrast design*. Some of these features cannot be associated to a point or area in space, are ephemeral or cannot be measured. Especially the issue of measurement concerns several features and needs to be further addressed.

The complete list is published under an Open Database License (ODbL) and can be downloaded [4].

Discussion. We discuss here some of the issues and findings that we learned from conducting the literature review and creating the list of accessibility features for all disabilities.

Literature reviews in multidisciplinary fields. Our systematic review yielded fewer results than expected, probably due to the unique search performed. Another researcher in the field of HCI and accessibility, Christin Engel, confirmed in a discussion that a systematic literature review in this area is difficult. The problem most probably lies within the keywords search: as the research field is a multidisciplinary one, each discipline (in our case cartography, architecture, IT, HCI) probably uses different terminology for the same concept. We therefore recommend adapting keywords to each database and performing a complementary search in the best conference proceedings and journals in the different fields.

Hierarchical representation of data. Our features include concepts that recur at different positions and levels in the hierarchy, with partially different sub-properties. One example are signposts, which are scattered over 11 features. This shows that a hierarchical representation is not best suitable for our data. A better representation could be a flat or object-oriented one with “has” or “is” relationships, like in A11yJSON[5].

Profiles. Our categories can be used to filter profiles in future mobile applications for indoor orientation and navigation and tailor them to users’ needs. Thus, they should not be excluded, even though the data should be structured differently. A combination of a flat representation and category assignments could offer most advantages.

Accessibility features and OpenStreetMap. Since we wanted to know which accessibility features can be mapped in OpenStreetMap (OSM) to be used in digital applications, we aligned the data to OSM keys and tags. We did this by introducing all features and the OSM keys and tags in a MySQL database - see Figure 3.1 for the database schema, and Figure 3.2 for the database statistics. We found that dimensions (length, width, height), as well as another 16.3 % of the features can already be mapped in OSM. Some of them can be defined by using multiple already existent OSM keys and tags. For instance, glass door can be described by *door=yes* and *material=glass*. Or, *maneuvering clearance in*

---

[4] `https://services.access.kit.edu/accessibility-features`

[5] `https://a11yjson.org/` Retrieved on 23 June 2021

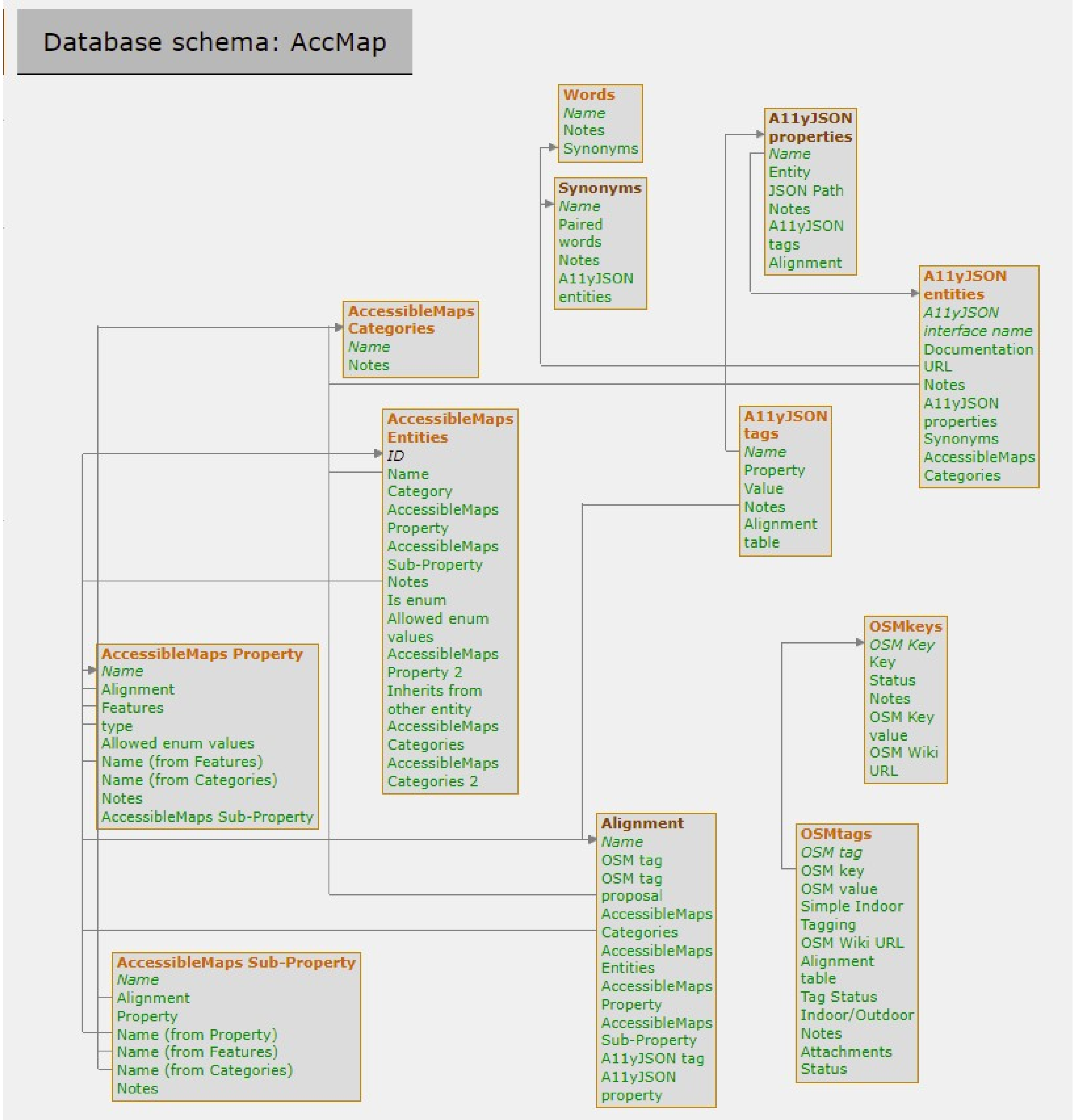
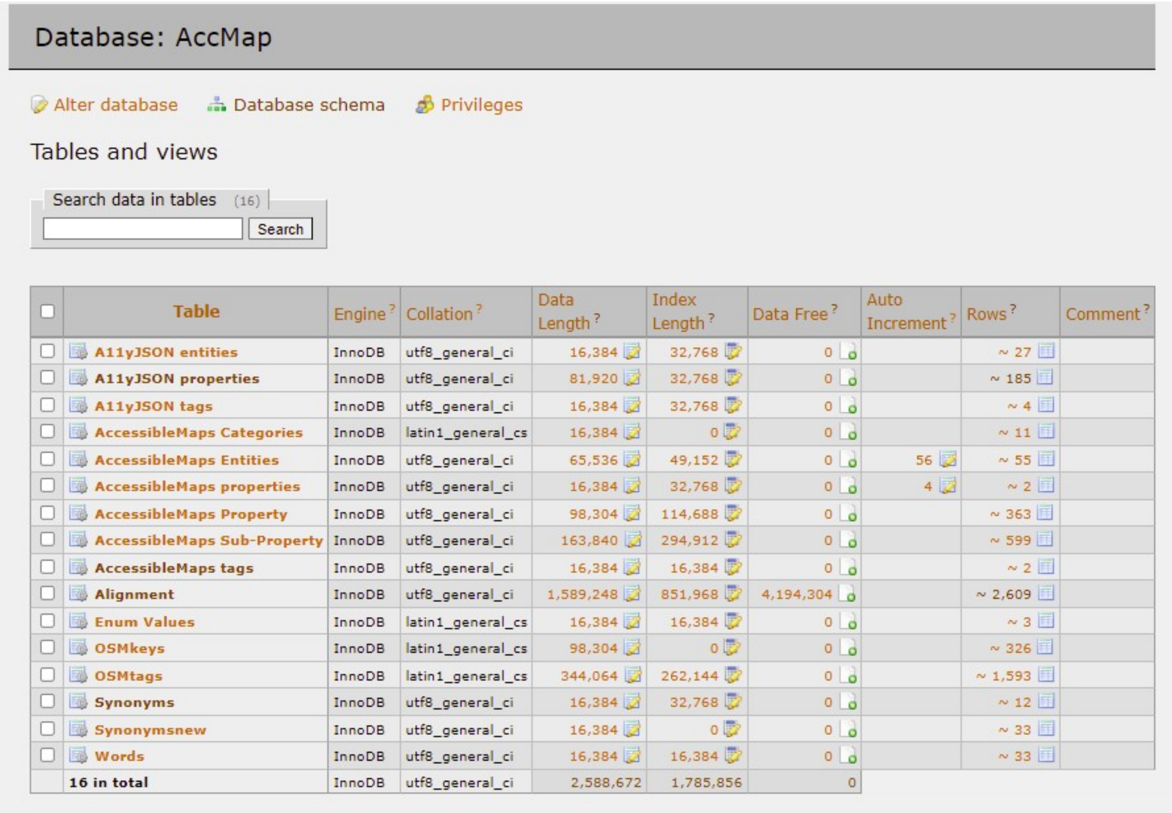


Figure 3.1.: The database schema of the database containing the accessibility features for all disability types and the alignment with the OSM keys and tags.

front of the ramp can be expressed by *ramp=yes* and the space in front of the ramp by *width=x* and *length=y*. However, according to the OSM rules, proposals to the OSM community still have to be made for using the existing tags and keys for mapping the objects. This ensures that everyone uses the tags and keys in the same way, which is essential for compatibility. For many features in our list, however, new OSM tags and keys need to be created from scratch. One example is the reception or information desk, which is extremely important for people with visual impairments for getting help on site. Another problem are the features that cannot be associated to a point or area in space, are ephemeral or cannot be measured, which are also difficult to map in solutions such as OSM. Few of them, such as *opening hours*, can be mapped in the description of the building, for instance, but for most others, solutions still have to be found.



Database: AccMap

Tables and views

☐	Table	Engine ?	Collation ?	Data Length ?	Index Length ?	Data Free ?	Auto Increment ?	Rows ?	Comment ?
☐	A11yJSON entities	InnoDB	utf8_general_ci	16,384	32,768	0		~ 27	
☐	A11yJSON properties	InnoDB	utf8_general_ci	81,920	32,768	0		~ 185	
☐	A11yJSON tags	InnoDB	utf8_general_ci	16,384	32,768	0		~ 4	
☐	AccessibleMaps Categories	InnoDB	latin1_general_cs	16,384	0	0		~ 11	
☐	AccessibleMaps Entities	InnoDB	utf8_general_ci	65,536	49,152	0	56	~ 55	
☐	AccessibleMaps properties	InnoDB	utf8_general_ci	16,384	32,768	0	4	~ 2	
☐	AccessibleMaps Property	InnoDB	utf8_general_ci	98,304	114,688	0		~ 363	
☐	AccessibleMaps Sub-Property	InnoDB	utf8_general_ci	163,840	294,912	0		~ 599	
☐	AccessibleMaps tags	InnoDB	utf8_general_ci	16,384	16,384	0		~ 2	
☐	Alignment	InnoDB	utf8_general_ci	1,589,248	851,968	4,194,304		~ 2,609	
☐	Enum Values	InnoDB	latin1_general_cs	16,384	16,384	0		~ 3	
☐	OSMkeys	InnoDB	latin1_general_cs	98,304	0	0		~ 326	
☐	OSMtags	InnoDB	latin1_general_cs	344,064	262,144	0		~ 1,593	
☐	Synonyms	InnoDB	utf8_general_ci	16,384	32,768	0		~ 12	
☐	Synonymsnew	InnoDB	utf8_general_ci	16,384	0	0		~ 33	
☐	Words	InnoDB	utf8_general_ci	16,384	16,384	0		~ 33	
	16 in total	InnoDB	utf8_general_ci	2,588,672	1,785,856	0			

Figure 3.2.: Statistics for the database containing the accessibility features for all disability types and the alignment with the OSM keys and tags.

First step towards standardization of accessibility features. The nonprofit organization SozialHelden[6] are interested in building a user interface for our database to allow querying with query languages like SPARQL or Jekyll. This would create a platform where experts from all accessibility fields can discuss the features, remove what they consider improper and add new features. This is especially important for outdoor features, as our set is mainly focused on indoor ones, and even there, we do not make a claim for completeness. SozialHelden are further interested in the standardization of accessibility features - thus, our indoor list becomes the first step towards standardization.

3.1.2. Accessibility Features for People with Visual Impairments

Since our focus is on people with visual impairments only, and the list resulted from the literature review addresses all disability types, we decided to compile a second list, only for visual impairment. We also wanted to have the accessibility features sorted by their importance.

After analyzing all the 26 sources, we came to the conclusion that the list for visual impairment should be, to a great extent, adopted from the ATMAPS project[7]. They compiled, in a human-centered way, a rated list of 136 indoor accessibility features relevant to people with visual impairments (Charitakis and Papadopoulos 2017). This list was created initially by seven project partners from four countries

---

[6] `https://sozialhelden.de` Retrieved on 08 June 2023

[7] `https://www.atmaps.eu/index.php/english` Retrieved on 08 June 2023

Categories	**Features**	**Few sample properties**
General building information (15)	Formal information (15)	Address; Name; Opening hours; Levels; Guide dog permitted
General help for orientation (13)	Reference points (6)	Prominent place as starting point; Sound source; Olfactory; Highly populated zones
	Tactile signpost (1)	Location of tactile description
	Map (4)	Type: floor plan, tactile, scale model
	Textual description (2)	Building; Route
Facility Daily Needs (33)	Services (15)	Availability; Type: information, service point, cashier, queue management ticket machine
	Toilet / toilet compartment (18)	Access: restrictions, public, EURO key, gender, wheelchair; Emergency call button; Drinking water
Building characteristics (90)	Doorways (30)	Name; Access: opening hours, key holder; Type: sliding, automatic, revolving, glass; Handle
	Floor (pathway) (12)	Access restrictions; Material; Change of the surface or texture; Tactile paving; Dimension
	Wall (3)	Material: solid, glass, railings
	Signs (1)	High contrast
	Rooms / venues / offices (28)	Name; Address; Level; Purpose
	Windows (3)	Shutters, rolling blinds or blinds
	Balcony (1)	
	Domestic engineering (11)	Type: socket, air condition, heating burner, pipes, wires
	Interior colors (1)	Conspicuous colors
Movables / Room equipment (66)	Furniture (40)	Type: hanging, seating, standing, seating, dangerous, orientation point; Size; Material: glass
	Machine or device (13)	Type: computer, printer, refrigerator, oven, cash machine
	Water (5)	Type: hot, cold; Sink; Water dispenser
	Light (4)	Type: lamp, light bulb, source; Switch location
	Telephone (2)	Type: pay phone, phone booth
	Fire protection (2)	Fire extinguisher; Fire hose
Change Elevation (51)	Elevator (16)	Number; Type; Access restrictions; Opening hours; Side door open; Floor announcement; Floor display; Levels connected
	Escalators (10)	Direction of travel; Lanes; Handrail location; Incline
	Stairs / stairways (18)	Type / shape: u-shaped, spiral; Levels connected; Handrail; Number of stairs; Tactile paving
	Moving walkway (1)	
	Change in ground height (6)	Steps; Ramp; Slope; Uphill; Downhill
Warnings / Obstacles (18)	Dangerous area / point (1)	Tape: hot to burn oneself
	Obstacles (17)	Position: hanging, overhead; Movable; Signs; Size; Automatic bars; Contrasting colors

Table 3.4.: Indoor accessibility features for visually impaired people only. One can see the similarity with the larger list (see Table 3.3. The categories and features are again complete, while the "properties" column only contains a few sample items. In parentheses is the number of pairs "feature-property" contained in the category or number of properties contained in the feature.

(Germany, Greece, Turkey and Cyprus), based on previous literature and their own experience as experts in the field. Then, the initial list was refined in 4 focus groups with people with blindness, resulting in a list with 226 entries. Finally, the list was rated in a questionnaire with 115 participants with visual impairments ranging from severe impairment to blindness, with ages between 18 and 64, coming from the four countries mentioned above. The participants rated the significance of information on audio tactile maps, as well as the frequency of occurrence of information on audio tactile maps. An importance factor is defined and calculated by multiplying the significance with the frequency of occurrence.

However, the ATMaps list also comes with a few limitations: first, most users come from the 4 countries mentioned, most of them in the same part of the world (Europe and Turkey). Secondly and more importantly, the target user group was blind people only, and the ratings refer to the significance to and frequency of occurrence of information (features) on audio-tactile maps. Finally, the focus in ATMaps is on the most important features, probably because on tactile maps only little information can be displayed. Thus, details to the features, such as properties and sub-properties, are often missing. Thus, the results of the ATMaps project must be carefully interpreted and adapted, depending on the particular scope they are meant for.

We thus reorganized the ATMaps list as follows - see also Figure 3.3:

- Categories: One expert in accessibility who is himself blind commented each feature in the list. Based on these comments, a sighted expert categorized the features.
- Properties: to the ATMaps list of indoor accessibility features we added properties and sub-properties that we obtained from our literature survey. For example, the ATMaps list contained the item "Toilet or Bathroom for men, women or people with disabilities". We added the following properties and sub-properties obtained from our literature review:
  - "access": "public"; "access restrictions"
  - "equipment": "EURO locking system"; "emergency button"; "diaper changing table"; "drinking water"; "hand washing"; "paper supply"
  - "accessibility level": "wheelchair accessibility"; "ADA compliant"; "colostomy support"
  - "availability": "opening hours".

  In this way, we added properties and sub-properties to the other features in the ATMaps list, resulting into a new list of indoor accessibility features for people with visual impairments - see Table 3.4. This list is actually a subset of the larger list for all disability types presented in Table 3.3 and contains almost 300 indoor accessibility features and sub-features (from the approximately 820 of the large list).

As one can see, the categorization in the list for visual impairment only is slightly different than the one for all disability types. Take for instance the example of "fire safety", which in the list for all disabilities is in the "Safety" category, while here it is in the "Movables / Room equipment" category. This suggests that features can belong to multiple categories so a flat categorization is not ideal.

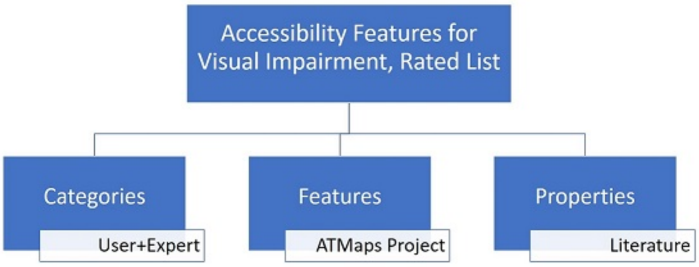


Figure 3.3.: The structure of the rated list of accessibility features for people with visual impairments, with categories, features and properties, and how these were obtained - whether from the systematic literature review, from the ATMAPS project or from user and expert ratings.

3.1.3. Conclusion

By means of a literature review, we identified the information needs of people with disabilities when travelling through unknown buildings. We compiled two lists, one for all disability types, containing more than 820 indoor accessibility features and sub-features (see Table 3.3), and a second list for people with visual impairments only, containing almost 300 features and sub-features (see Table 3.4).

We created a database for the large list, in which we aligned our features with OSM keys and tags. This way, we identified what can be already mapped in this open geographic database (world map) to be directly used in applications, and what is still to be done. Finally, together with the nonprofit organization Sozialhelden, we set the cornerstone for standardization of the features list. This in turn would greatly facilitate the development of applications that support people with disabilities in travelling independently.

3.2. Non-Functional Requirements for Assistive Output Interfaces

We have seen in the previous section that people with disabilities in general, and people with visual impairments in particular, need a large amount of information when travelling. The question that arises is *how* to convey this information to the users. We start by specifying the non-functional requirements for assistive interfaces in general and auditory displays in particular for people with visual impairments. Non-functional requirements are "features of the system that are not directly related to the actual services provided but relate to the manner in which those services must be provided. Some classical examples [...] are efficiency, reliability, timing and safety" (Dix et al. 2004). Thus, non-functional requirements specify *how* a system should be, as opposed to functional requirements, which specify *what* a system should do. Non-functional requirements can be applied to categories of systems, while functional requirements are specific to one particular system. Since we address mobile systems that support orientation and mobility in general, and not one particular system, we will consider here non-functional requirements only. The main question addressed by this section is: what are the most important requirements that an output interface of mobile assistive systems must meet, in order to be both usable by and useful to people with visual impairments when travelling independently?

The market for assistive systems for people with visual impairments is a rather small one. While there are many people with visual impairments worldwide, compared to the overall population, they are just a small number. Besides, the majority of them are older people who are less likely to use technological solutions. Finally, the user group is so heterogeneous, that one solution rarely fits more than a few customers. This is probably the reason why there are no standards related to the interfaces of assistive systems for this population. The closest it gets to a standard requirements definition is defined in the international standard ISO 9241-171:2008, which we also used as a starting point for defining our requirements. However, this standard applies to software in general, while we need requirements for the output interface (not the entire software) of mobile assistive systems for people with visual impairments.

Some research projects approach this issue and define functional and non-functional requirements. Miao et al. (2011), for instance, gathered from 6 blind people user requirements - both functional and non-functional - for a mobile application meant to assist in the independent navigation through a public building. Alkhanifer and Ludi (2014) carried out several studies consisting of semi-structured phone interviews with 24 people with visual impairments from various English speaking countries around the world and 6 orientation and mobility instructors, followed by a large online survey with 65 English speaking blind people. They gathered various user needs when navigating unfamiliar buildings and propose a set of requirements for assistive indoor navigation systems. They also point out cultural differences, for instance in terms of Braille usage or the extent of information needs inside buildings based on average building sizes in the different countries/continents. Their focus, however, is restricted on *open* indoor spaces. Hersh (2018) provides "a brief overview [of how blind people travel] rather than an indepth consideration of all the important issues". She draws a set of general requirements that interfaces for blind people should meet, such as: high quality, low cost, easy to use while the user is moving, non-interference with environmental sounds, a choice of (vibro-)tactile and audio output.

Except for Hersh (2018), all these projects inspired our work in Section 3.2, but in themselves did not solve our problem, as they have a different scope than we do, namely they do not address non-functional requirements, they aim at mobile applications, not their interfaces or they only address indoor travel. Only the work in Hersh (2018) meets precisely our scope. Unfortunately, the book was published after we conducted our research, so we did not use it. However, it seems that our requirements are in line with their results.

In our approach, we first collected requirements in an initial literature review. Then, in an online questionnaire, we asked experts in the field to give us feedback on the obtained requirements. Finally, we refined our list based on the experts' feedback and a second literature review. The requirements were rated by experts according to their importance. Thus, in the end, we obtained a final list of 13 requirements with criteria, sorted by importance.

Similar to some standards, our requirements "can be utilized throughout the design process (e.g. as specification and guidance for designers during design or as a basis for heuristic evaluation)" (ISO/TS 9241-126:2019).

3.2.1. Initial Set of Requirements

We identified a set of requirements in an iterative way. In a first step, we carried out an initial literature review, in which we included highly relevant sources known to us, such as results from the requirements analysis of the Terrain project[8] (performed earlier by other project contributors), other known and similar projects such as InMoBS (Vollrath et al. 2015) and sources found with Google and Google Scholar searches. The sources thus gathered included relevant information ranging from high-level such as the Universal Design Principles, short UDP (Connell et al. 1997), to more fine grained information, available in international and local standards related to accessibility of software and technical products (ISO 9241-171:2008, DIN-Fachbericht 124) as well as research in the area of navigation for people with visual impairments (Dakopoulos and Bourbakis 2010, Gallagher et al. 2014, Kristjansson et al. 2016) and wearable technology (Zeagler 2017).

Requirements were selected based on the following inclusion / exclusion criteria:

---

[8] `https://www.access.kit.edu/english/1248.php` Retrieved 08 June 2023

- Requirements must relate to mobile systems, i.e. the interface should not require a fixed workplace or a heavy hardware setup.
- Requirements must address non-visual displays. We did not consider requirements related to colors, fonts, size of text and images, or anything else visual.
- Requirements must address people with visual impairments. We did not consider other impairment types such as hearing, mobility, cognitive, etc.
- Requirements must be related to travel scenarios, both indoors and outdoors. We did not consider other scenarios, such as surfing the Internet or doing everyday tasks in familiar environments.

At the end of this initial literature review, we obtained 14 requirements together with a preliminary set of sample criteria. For instance, for the requirement “comfort” (Kristjansson et al. 2016), two sample criteria were: “output devices should be lightweight” (DIN-Fachbericht 124) and “they should leave the user’s hands free” (Kristjansson et al. 2016). The entire list of requirements, together with sample criteria and, where needed, examples, can be found in Table 3.5.

3.2.2. Expert Assessment

In a second step, we asked the opinion of 22 experts on the requirements gathered, in accordance with usability best practices, which recommends evaluating early.

In 2009, Cerpa and Verner (2009) analyzed 70 failed software projects and found out that in 73% of cases, one of the failure factors was “inadequate requirements”. Another work, that of Lane et al. (2016), makes a similar statement on the importance of requirements, namely that “any issues with the elicited requirements have an impact on the project as a whole and in some cases can lead to project failure”. In order to avoid this problem, we decided to evaluate our requirements early. According to Dix et al. (2004, p. 319), “evaluation should occur throughout the design life cycle, with the results of the evaluation feeding back into modifications to the design”. Thus, evaluation should not be carried out at the end of the project only, when feedback can hardly be considered anymore. The earlier one starts evaluating, the sooner can problems be identified and corrected. We thus wanted to assess our requirements before identifying long lists of criteria for each requirement.

One practical and economical way of evaluating early during the design is through “expert reviews”. This is a far more effective way than the commonly used strategy of asking colleagues or customers (Shneiderman et al. 2018, p. 171), and more practical than evaluating with users. As people with visual impairments are very diverse in their capabilities and preferences, asking 10 different users the same question would probably get us 10 different answers, so we would have to ask a large number of end users in order to draw conclusions. An expert evaluation with a smaller number of experts would be more practical. Although there are researches that found user evaluations better, in some respects, than expert evaluations (Yen and Bakken 2009), these have the limitation that their experts are HCI experts, while the users come from a different field, such as medical. Thus, experts do not have an understanding of the users’ needs. Who the experts in our case are, and why they are appropriate for our task, we explain below.

#	Requirement	Weight	Sample criteria (guidelines)	Example	Sample references
R1	Safety	5.0	The user output should not replace, dominate or interfere with safety-related information in the environment.	In-ear headphones are not safe for people with visual impairments when traveling.	Proposed by experts
R2	Accessibility	4.8	The user output should be accessible to people with various and multiple disabilities and should address at least two of the three senses (sight, hearing, touch).	A device with voice output alone is not accessible to people with deafness. The deaf-blind population should be considered (E18).	Connell et al. (1997)
R3	Adjustability	4.8	The user should be able to adjust the output according to their preferences.	The user should be able to adjust: the speech rate and volume of voice output, choose between computer and human voice (E4) or between voice and tactile output (E21); set the sensitivity for obstacle detectors to avoid false alarms (E4).	European Telecommunications Standards Institute (2016)
R4	Simplicity	4.7	Unnecessary complexity should be avoided.	Voice output should use clear and simple language.	Kristjansson et al. (2016)
R5	Consistency	4.6	In similar situations, information should be conveyed to the user in a similar manner.		European Telecommunications Standards Institute (2016)
R6	Clarity	4.6	Information should be clearly conveyed to the user. There should be sufficient contrast between the interface and the environment.	The volume of voice output should be at least 10 dB above ambient noise, but not exceed a given limit.	DIN-Fachbericht 124
	~~Minimization of cognitive load~~	4.6	The user should not be cognitively overloaded by the user output.	Mental demand should not be significantly higher than for the same task without the system.	Gallagher et al. (2014)
R7	Comfort	4.6	Output devices should be lightweight, hands free, not extend beyond the user's self-perceived height or restrict the user's movement, should not become too warm.		Kristjansson et al. (2016), DIN-Fachbericht 124, Zeagler (2017), Connell et al. (1997)
R8	Error tolerance	4.5	Prevent users from making errors. Help users recover from errors.	A navigation system should not say "turn now" when the location accuracy is over 10 meters. Hints should be accurate and not make mistakes.	Connell et al. (1997), ISO/IEC Joint Technical Advisory Group (2014)
	~~Reachability~~	4.5	All elements of the output devices should be easily accessible both while sitting and standing. Different hand and grip sizes should be accommodated.		Connell et al. (1997), Zeagler (2017)
R9	Speed	4.4	How quickly can information be delivered to the user? Time critical feedback should be very fast.	Warnings of an obstacle that appears just ahead of a blind user should be transmitted instantly.	European Telecommunications Standards Institute (2016)
R10	Social acceptability	4.3	A system and its user interface should be aesthetically pleasing and not stigmatize the user.	A helmet with headphones sticking out is not socially acceptable.	Connell et al. (1997), Kristjansson et al. (2016)
R11	Intuitiveness	4.2	The user should be able to use the system with little or no training.	Give directions as clock-face instead of cardinal points.	Kristjansson et al. (2016)
R12	Compatibility	4.0	The output should be compatible with different output devices, including other supporting technologies.	3D sound cannot be played on bone conducting headphones, so it is not compatible.	Connell et al. (1997), ISO/IEC Joint Technical Advisory Group (2014)
R13	Affordability / cost efficiency	3.7	The interface should be affordable to most people, including those in developing countries or the solution should be financed by a cost bearer.	Mono sound can be played on any headphones and does not need special, expensive devices.	Dakopoulos and Bourbakis (2010)

Table 3.5.: The 14 initial requirements with average weights given by the experts, together with a few sample criteria, examples and some references. The violet-colored items were suggested by the experts during the evaluation - this includes the "safety" requirement. The two requirements in ~~strikethrough~~ font were removed from the final list.

When a person's sight deteriorates so much that independent living is hardly or not at all possible anymore, that person must go through a rehabilitation process and learn how to do things differently, compensating for the missing sight. One of the most important parts of this rehabilitation process, besides a training in practical life skills, is the orientation and mobility (O&M) training. Orientation and mobility trainers teach people with visual impairments how to travel again independently using a white cane. Often they also teach the use of other tools such as a guide dog or electronic aids. Since the training is always very individual, mobility trainers must first get to know their customers very well with respect to their mobility, orientation and technology background and skills. As such, the trainers have a good overview of the needs and preferences of all their customers. In countries such as Germany, where the health insurance system is funding the rehabilitation, mobility trainers often have many customers. Thus, they have an overview over the needs of *many* people with visual impairments with respect to *mobility*, *orientation* and related *electronic aids* - which is exactly the expertise that we need in order to evaluate our requirements. Thus, instead of asking a large number of people with visual impairments, we ask a smaller number of mobility trainers. End users should still be involved in the development and evaluation of assistive systems, but for this stage of development, and given the particularity of our target user group, testing with experts is the better option.

Method. We thus prepared an online questionnaire and sent personalized invitations per email to 181 experts in various fields related to visual impairments from Germany, the vast majority of them being mobility trainers. Most email addresses were taken from the website of the German Federal Association of Rehabilitation Teachers for the Blind and Visually Impaired e.V.[9], while few others were personal contacts of ACCESS@KIT. The purpose of the questionnaire was to remove unnecessary requirements, to identify further relevant ones and to obtain an importance rating. A total of 22 German mobility trainers answered the questionnaire. We will refer to them as *E1* through *E22*, where the letter *E* stands for "expert".

In the questionnaire, we presented the initial set of 14 requirements, together with a few examples from the sample set of criteria. Table 3.5 lists the 14 requirements as well as most of the sample criteria and examples that were used in the questionnaire. The participants were asked to rate the importance of each requirement on a Likert scale from 1 (not at all important) to 5 (extremely important). At the end, they were asked to name further requirements that they consider important and to assign them a rating.

Results. The weights from all experts were averaged - see Table 3.5, third column. All requirements except for one were considered important or very important, having an average rating of 4.0 or more. The maximum standard deviation was 0.98. The weight 1 (not at all important) was never assigned, and 2 (not important) was assigned 4 times to these 13 requirements. The least important requirement was found to be *affordability*, with an average rating of 3.8 and a maximum standard deviation of 0.98. This could be because in Germany, the costs of assistive technology (AT) can be financed from various official funds, so users do not have to pay for them. We suspect that this does not apply to other countries, where the cost factor might play a much more important role.

Eight mobility trainers answered the part of the questionnaire which asked to name further requirements. Two of them made suggestions related to *safety*: E2: "*The output should not obscure safety-related information in road traffic (e.g. in-ear headphones)*" and E17: "*The user output of a mobility assistance system should not replace or dominate or interfere with immediate perceptions*". Thus, *safety* was identified as an extra requirement. It was added to the list with a rating of 5.0, as suggested by expert E2, due to its importance in navigation scenarios.

Six mobility trainers made comments which we included as criteria or as examples to already existent requirements - see the violet text in Table 3.5.

---

[9] `https://www.rehalehrer.de/rehabilitationslehrerinnen/orientierung-mobilitaet/suche` Retrieved on 08 June 2023

3.2.3. Refinement of Requirements

Based on the expert feedback, we refined the list of criteria by further searching in the literature. Finally, we restructured the requirements such that all the example criteria found fits well into the requirements, and computed for each requirement an importance score, according to the expert feedback.

We further analyzed several European and international standards and recommendations, based on expert feedback, and also in order to identify further example criteria. We reviewed relevant literature, namely conference papers from the two most recent editions at the time this research was conducted: CHI[10] 2018 and 2017, ASSETS[11] 2018 and 2017 and ICCHP[12] 2018 and 2016, as well as other relevant research papers and standards found using Google and Google Scholar. In the end, the following sources have been found and used to extend the existent criteria: ISO/IEC Joint Technical Advisory Group (2014), European Telecommunications Standards Institute (2016), Technical Committee ISO/TC 159, Ergonomics, Subcommittee SC 4, Ergonomics of human–system interaction (2017), Kirkpatrick et al. (2018), Richens et al. (2019), Funka Nu AB (n.d.).

While gathering further criteria, we also restructured the requirements to better fit the new criteria. For instance, the requirement "Minimization of cognitive load" was removed, as it relates to the user, not directly to the interface itself. In order to minimize the cognitive load, the interface must be designed taking into account the limitations of the user's memory. Examples of criteria that relate to the interface are: "segment and sequence complex material" (Spencer 2015) or "draw on users' existing knowledge and mental models" (DiMaggio 1997). But this criteria can be included in other, already existent, requirements: "Simplicity", "Intuitiveness", "Consistency" and "Clarity". Or, otherwise said, interfaces that are simple, intuitive, consistent and clear minimize the cognitive load of the user. Thus, the requirement "minimization of cognitive load" is not needed anymore and can be removed. In a similar way, the requirement "Reachability" was integrated into "Comfort". The reasoning for integrating the two was the following: a comfortable interface must have easily reachable elements, or the other way around: if the elements of a mobile interface are not easily reachable, then using the system is not comfortable.

Finally, we computed an importance rating based on the expert evaluation results by averaging the ratings given by the mobility trainers.

Discussion and limitations. On top of the requirements list, thus slightly more important than the others are safety, accessibility and adjustability. While the first two naturally emerge from the travel context and the target user group, i.e. people with visual impairments, the third requirement in the top three list suggests that designing for the average user will not work well for this particularly heterogeneous user group. Instead, one must offer as many options as possible, together with the possibility to configure the system according to individual needs.

---

[10] The ACM CHI (pronounced "kai") Conference on Human Factors in Computing Systems is an A-rank conference, the most important one in the field of Human-Computer Interaction (HCI). It includes a track on accessibility.

[11] The International ACM SIGACCESS Conference on Computers and Accessibility is an A-rank conference, and probably the most important in its field.

[12] The biannual International Conference on Computers Helping People with Special Needs is an important conference in the area of accessibility in Europe.

One limitation of the study is that the experts all came from Germany, which introduces a bias towards the German society. For other countries, the importance ratings might not apply. Another limitation of our study lies in the restructuring done to the requirements after the evaluation, including the addition of new requirements and criteria. This could lead to errors in the importance ratings. However, the most significant aspect is that all requirements were found to be important. This suggests that a lot of factors must be taken into account when designing an interface in this context and for this user group.

3.2.4. Conclusion

Through extensive literature research and an expert study with 22 mobility trainers from Germany, we defined a set of 13 non-functional requirements for the output interface of mobile assistive systems for people with visual impairments. The experts also assessed the importance of the requirements. Except for affordability, which was rated on average below 4.0 on a scale from 1 (not at all important) to 5 (very important), all other requirements were rated as important (average over 4.0). The experts also suggested an extra requirement, namely *safety*, which was added to our list. We argue that in our case, testing with experts replaces the testing with many users, as our experts know very well the users in the context of travel - thus, they are experts in the domain being tested. Given the heterogeneity of the target user group and the difficulties in finding participants for evaluations, testing with experts is more efficient in this particular case.

4. Speech Auditory Displays

Speech is currently the most widely used alternative interface to visual displays, especially by people with visual impairments. This chapter's focus is on speech output interfaces for people with visual impairments in the context of travel. The question addressed in this chapter is: *how* to convey information to people with visual impairments during travel *using text-to-speech*?

As seen in section 3.1, the list with information - accessibility features - needed by people with visual impairments when travelling is very large. At the same time, especially in the case of people with blindness, the users rely much more on their hearing in order to stay safe and find their way, compared to sighted people. Thus, the sense of hearing must not be overwhelmed with too much information coming from an additional source, namely an assistive system. The speech interface must be carefully designed and comply with the non-functional requirements introduced in section 3.2. Moreover, when traveling with little or no sight, one must differentiate between outdoor and indoor travel. Outdoors, the space is broad and it is difficult to get an overview of the entire area between the starting point and the destination of a route. Accessibility features or "objects" that are relevant to people with visual impairments are more sparse outdoors than indoors. While outdoors navigation instructions are needed, indoors, describing the entire building or parts of it is feasible and helps users create a better mental map of their environment. This way, users can make better and more spontaneous decisions about their route. In the next sections, we will analyze these two contexts, indoors and outdoors, separately, while trying to answer the question mentioned above, namely: *how to convey information through speech?*

In section 4.1 we focus on text-to-speech interfaces for *navigation* enhanced with *object* announcements. In section 4.2 we focus on text-to-speech interfaces with information about accessibility features inside of buildings that can be used for orientation, searching and obstacle avoidance.

4.1. Speech Auditory Display for Outdoor Travel: Enhanced Navigation Information and Objects Announcements

In this section we address the question of *how* to convey information to people with visual impairments during *outdoor* travel.

Since the introduction of the global positioning system (GPS), navigation applications have found wide acceptance. However, the speech feedback of most of these applications was designed for use in cars (Correia et al. 2002) or, at best, for sighted pedestrians (Allen 2000). Thus, information that is critical for people with visual impairments, such as the roadside on which one should walk, when to cross, or the exact position and form of the crosswalk - all this is missing in mainstream navigation applications. Besides routing instructions, people with visual impairments also need information about objects in the environment, whether landmarks that "inform their decisions during navigation" (Sato et al. 2019a), points of interest or obstacles. Recent advances in smartphone technology and computer vision make it possible to retrieve ever larger amounts of data from the environment, and to communicate them to the user. How to do this in a way that is appropriate for people with visual impairments, i.e. without overwhelming the users and posing a safety risk, is described here. In the following part, we describe our grammar, which we designed to communicate to the user routing instructions as well as information about the environment in a structured way. It is very important for a speech-based interface to be tailored to the local community (in our case the German one) to ensure its quality, as a user of a mobile assistive app in the travel context stated: "This app has become worse, not better. If you set the language to German, there are a lot of translation errors. Unfortunately, many navigation apps designed specifically for the blind and visually impaired are often in the same poor quality as this one. Sad" (Baake 2020). Our grammar is tailored for the local community, i.e. for the German language.

4.1.1. Creation of the Grammar through an Expert Round-Table

This subsection is based on our ASSETS publication Constantinescu et al. (2019).

A literature search on existing voice announcements of navigation systems gave only few results. Often examples are mentioned (Gaunet and Briffault 2005, Kulyukin et al. 2008, Pressl and Wieser 2006), but the underlying grammar is not explained and it is unclear whether a grammar was at all specified. In Gaunet and Briffault (2005), for instance, route descriptions produced by blind pedestrians were analyzed in order to produce functional specifications for a navigational aid. The work is restricted to 'simple and structured' urban areas. Balata et al. (2018) propose a set of navigation instructions that take roadside into account. Their proposal concerns the transitions between indoor and outdoor environments and the usage of public transport. The instructions are verbose and are implemented using templates. Nicolau et al. (2009a,b) describe approaches to a grammar used for indoor purposes. They analyzed how people with visual impairments verbalize routes. However, these descriptions are difficult to generalize and therefore cannot serve as a basis for our grammar. We also interviewed other researchers involved in similar projects, such as the Mobility project (Spindler et al. 2012). However, this did not get us much further either, as their main goal was to provide descriptions of the environment in indoor environments, while our main goal is to provide enhanced navigation instructions and environment (objects) information for *outdoors*. Thus, we decided to create a new grammar for this task.

An initial analysis of nine navigation systems, including the car navigation systems from Ford and Chevrolet (Novartis 2015, Transition Technologies S.A. 2018, Swiss Federation of the Blind 2018, Google 2018, Apple Inc. 2018, Spindler et al. 2012, komoot GmbH 2022) performed by Vanessa Petrausch, resulted in the following two syntaxes, written as a regular expressions (Stubblebine 2003):

Syntax 1:

$$Distance\ Object^{*}\ Direction\ Action\ Street? \tag{4.1}$$

Syntax 2:

$$Distance\ Direction\ Object^{*}\ Action\ Street? \tag{4.2}$$

where “*” matches zero or more instances, and “?” matches zero or one instances. Here is an example of instructions using syntax 1 (4.1):

German:	In 35 metern links abbiegen auf Hauptstraße.
Literal translation:	In 35 meters left turn on Main Street.
English:	In 35 meters turn left on Main Street.

Syntax 1 (4.1) follows the natural order in which actions are performed. For instance, taking the example above: the user first walks *35 meters*, then makes a *left turn* and continues on *Main Street*. According to Allen (2000), this should “lead to more successful wayfinding efforts than will route directions that violate natural temporospatial order”.

In order to assess and refine the initial syntaxes and clarify interface aspects that are specific to people with visual impairments (what to transmit when), we conducted a round-table with 5 experts: two mobility trainers, an accessibility expert, a linguist and an assistive technology adviser who is himself blind. The participating experts had at least five years of experience with the target population at the time the round-table took place and could give informed opinions on topics related to interfaces and language in the travel context. The round-table was organized as a semi-structured interview. The questions that were prepared in advance were derived from a previous requirements analysis conducted within the Terrain project[1] by other researchers. We asked those questions that were left unclear or unanswered in the requirements analysis. We thus asked the experts to give us feedback on the two syntaxes (4.1 and 4.2) and answer questions related to street sides, walking direction, crosswalks and other accessibility features and when to announce them. Further questions and insights arose during the round-table. One of our main goals was to find out how to reach a suitable level of verbosity of the announcements, without neglecting any important information. We present here the most important findings.

Syntax. Our grammar includes sentence elements that are not encountered in conventional pedestrian navigation systems. Here is an example of a sentence using syntax 1 (4.1):

German:	In 15 Metern an Ampel links überqueren von Hauptstraße.
Literal English translation:	In 15 meters, at traffic lights, left cross Main Street.

The expression “left cross” means that the user needs to turn left ($-90°$) and then cross the street.

The additional information provided by us resulted in both proposed syntaxes sounding unnatural. Thus, the experts agreed upon a syntax that works well in all situations, independent of the amount of additional information provided, according to a natural word order in German (see 4.3). The resulting syntax is based on syntax 1 (4.1), with the information about the street name moved from the end of the sentence more upfront, right after distance. The information about roadside, which was nonexistent in the preliminary syntax, being an information specific to people with visual impairments,

---

[1] https://www.access.kit.edu/1248.php

was inserted after the street name. We provide here the syntax as proposed by the German experts, together with the English translation, which takes into account the English natural word order. Please note that the English version of the syntax is only presented here for exemplification and for easing the understanding for English readers, and is not intended to be used in system implementations.

German syntax proposed by experts:

$$Distance\ Street?\ Roadside?\ Object^{*}\ Direction?\ Action \tag{4.3}$$

English equivalent of the syntax proposed by experts:

$$Distance\ Action\ Direction?\ Street?\ Roadside?\ Object^{*} \tag{4.4}$$

Objects (accessibility features) appear in our grammar both within the navigation instructions (e.g. "cross left at *traffic lights*") and separately, in independent announcements about things in the environment (e.g. "*park bench* at 3 o'clock in 20 meters"). Whether the object is included in the navigation syntax or announced separately, depends on several factors:

(1) The type of object. Only objects that can be used as landmarks by people with visual impairments are included in navigation announcements. A large park, for instance, cannot be used as a landmark, as it is not precise enough;

(2) The distance between the object and an action point in a navigation instruction. Only nearby objects (within 25 meters) are included in the navigation instructions;

(3) Users' configuration. Users may include or exclude objects to be announced. However, a differentiation should be made between objects for navigation and those for independent announcements. That is, a user might choose to include a specific object in the navigation instructions, but to exclude it from the individual announcements, or the other way around.

The independent objects announcements are made with a shortened and modified version of the syntax, where direction comes before distance. In the case of *navigation instructions*, the distance must be announced first, so the users know how much time they have to react. It is also the natural order in which actions are performed. In the *object announcements*, the most important information is the type of object, as the users first decide whether the object is important at this point in time or not. If the object is not important, the rest of the message can be ignored. Otherwise, the object's location might also be needed. Here, the direction comes first, as the users would first turn into the object's *direction* and then walk the *distance* to the object if they wanted to reach it. The final syntax for announcing objects is - this time the same for English and for German:

$$Object\ Direction\ Distance \tag{4.5}$$

Direction specification. The direction of objects in the *independent objects announcements* is conveyed using the clock-face system or clock position as used by pilots (Mariner 2007). This technique is now commonly used by Orientation and Mobility (O&M) instructors (Hersh 2018) and in assistive technologies for people with visual impairments that assist in navigation (Zeng et al. 2017, Alkhanifer and Ludi 2014, Sánchez and de la Torre 2010): twelve o'clock is in front of the user, three o'clock to the right, six o'clock behind and nine o'clock to the left. The direction in *navigation instructions*, on the other hand, is conveyed using *slight/sharp left/right* (German: "leicht/scharf links/rechts"). The reason for this difference in conveying directions is first of all, to ensure a similarity to already existing solutions used by people with visual impairments such as Google Maps or BlindSquare. Secondly, directions expressed with "left-right" are more natural and easier to understand, while clock directions are more precise. Object announcements must be more precise, for instance, when the user is searching for an object. Navigation instructions on the other hand can be less precise, since blind users orient themselves by following the edge of the sidewalk, whether curb, wall, fence or grass.

Driveways. According to the experts, driveways don't have to be mentioned, except those from gas stations and parking garages, which are usually wide and very dangerous. For these two particular cases (gas stations and parking garages), the instruction is spoken only once, 150 meters (500 feet) in advance. This is because these objects are large and are usually represented in mapping solutions using a single point in the middle of the building or of the property, which does not clearly show where the driveway is situated. In comparison, all other accessibility features are announced only up to 25 meters in advance approximately (subject to localization accuracy).

Special objects. According to the experts, objects such as stairs, ramps and any kind of barriers such as poles, railings to slow down bicycles, gates, fences, turnstiles, do not have to be mentioned, as they can be detected with the white cane. They could be included, however, in some advanced options, such that the user may choose to have any of them announced. Rails must be announced in advance, but only if the user is supposed to cross them. Start and destination of the route also need to be mentioned.

Intersections. Intersections should be explained upon user request, except on unknown routes, where they should be explained automatically. There should be an option to allow switching this feature off and on again. The system should convey information about the form of the intersection (X, T, Y, star, rotary) and the direction of the streets in relation to the user.

Deviations from the route. The user should have the choice between reconfiguring the route or continuing with the old route. The user should have the option to turn automatic reconfiguration of the route on or off again from the settings. By default, the system should not reconfigure the route automatically without at least announcing the user, as this may lead to confusion and loss of orientation.

Orientation / walking direction. At the beginning of a route, the system should tell the user in which direction to start walking. The instructions could be spoken and should include the cardinal direction (compass points: North, East, etc.). Moreover, the user might be informed about the correct direction by vibrations of the smartphone, like in ViaOpta Nav (Novartis 2015). For that, the user holds the smartphone parallel to the ground, pointing forwards, and stands facing the street, as for crossing. Then the user turns until the phone vibrates to signal the correct direction, which should be left or right from the user's original position. Using this method, one can determine the starting position more reliably, and errors due to device response time are avoided. The users should be able to use this method anytime, upon request, in order to orient themselves throughout the travel route.

Roadside. Roadside only needs to be announced on start, destination and before driveways of gas stations and parking garages. For the rest of the route, one can orient using building walls, grass or the sidewalk border.

Distance to an event. The experts agreed that when using GPS alone for positioning, no directions should be given in less than 10 meters before the event, because the GPS accuracy combined with user's walking speed could result in the instruction being spoken after the user passed the event, which can be very confusing. As a note, blind pedestrians, especially young ones who were born blind, can walk very fast.

Curves. When there is no intersection nearby (the user walks on a simple road segment), and the angle is ample, then the curve does not need to be announced. Announcing such a curve may in fact confuse the user.

Changing the roadside. Changing the roadside by crossing through the middle of an unmarked and unsignalized street (instead of at the intersection at the end of the street) may be permitted in Germany, depending on traffic (Bundesministerium für Verkehr, Bau und Stadtentwicklung und Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit 2013), and is sometimes recommended by mobility trainers for people with visual impairments. In such a case, the system should warn that the street is to be crossed somewhere in the middle, where it is less dangerous, because there are no markings or traffic lights at the end of the street. No distance should be specified, as the users must decide by themselves where to cross depending on traffic, parked cars and other local characteristics.

Street name changes. When the name of the street on which a user currently walks changes, but there is no event coming up (intersection, turn), then the name change does not have to be announced.

Verbosity and configurability of the interface. Less information is better. However, the user should be able to request or turn on information from the options.

Language. Use plain language. Not everyone understands technical terms such as *perpendicular, parallel, vertical.*

User background. The user must have passed Orientation and Mobility training and should walk independently on *known routes* even without an assistive system. The user should never rely on a system alone, as this would be too dangerous.

Here are some examples of *default* speech instructions, for both navigation and object announcement, generated with the final syntaxes (4.3, 4.4, 4.5) and considering the results of the expert round-table. *Default* means that the users should have the possibility to add, remove or change the order of words in the sentences according to their own preferences:

German:	In 15 Metern Hauptstraße an Ampel links überqueren.
	In 10 Metern scharf rechts abbiegen, dann 75 Meter Hauptstraße folgen.
	Parkbank auf halb zwei in 11 Metern.
Literal English translation:	In 15 meters Main Street at traffic lights left cross.
	In 10 meters sharp right turn, then 75 meters Main street follow.
	Park bench at one thirty in 11 meters.
Proper English translation:	In 15 meters cross left Main Street at traffic lights.
	In 10 meters turn sharp right, then follow Main Street for 75 meters.
	Park bench at one thirty in 11 meters.

The complete grammar specification can be found in Appendix A.1. Please note, however, that the grammar in the appendix includes improvements suggested during the evaluation (see subsection 4.1.3 below).

In order to assess our grammar, we implemented it into a prototypical application and conducted an on-field user study in the city with people with visual impairments. Please note that the implementation is only exemplary, and does not include all the details and use cases covered by the grammar.

4.1.2. Prototypical Implementation of the Outdoor Grammar

We implemented the grammar into an application initially programmed by the project partners in the Terrain project, the company iXpoint Informationssysteme GmbH together with researchers from KIT, led by Daniel Koester. Our goal was to analyze whether additional information at signalized crosswalks conveyed through a well designed speech interface can support people with severe visual impairments in unfamiliar environments.

The built environment is rarely accessible for people with visual impairments and especially for blind people. Tactile paving, acoustic signals in traffic lights, information panels in Braille, acoustic information in buses and bus stations - these and many other such facilities help people with visual impairments travel more safely. However, they are implemented only scarcely, due to a lack of awareness over their importance and more so due to their high costs. Crosswalks in particular pose a major safety threat for people with visual impairments (Williams et al. 2013). In a study with 48 participants during 416 crossings, Barlow et al. (2005) found that visually impaired pedestrians started crossing during the "walk" (green) phase of the traffic lights in only 48.6% of the cases, and they completed the crossing after the onset of perpendicular traffic in 26.9% of cases. Common problems faced by people with visual impairments when crossing the street are: (i) finding the crosswalk and the correct position to cross, (ii) properly aligning for the crossing, (iii) finding the pushbutton of a pedestrian traffic light - if there is one available, (iv) deducing the walking phase of an inaccessible traffic light, (v) finding the correct time to cross and (vi) keeping the correct direction while crossing the street when relying on hearing alone (Liao 2012, Ross and Blasch 2000, Hassan 2012). In familiar environments, most of these problems can be compensated through learning and *Orientation*

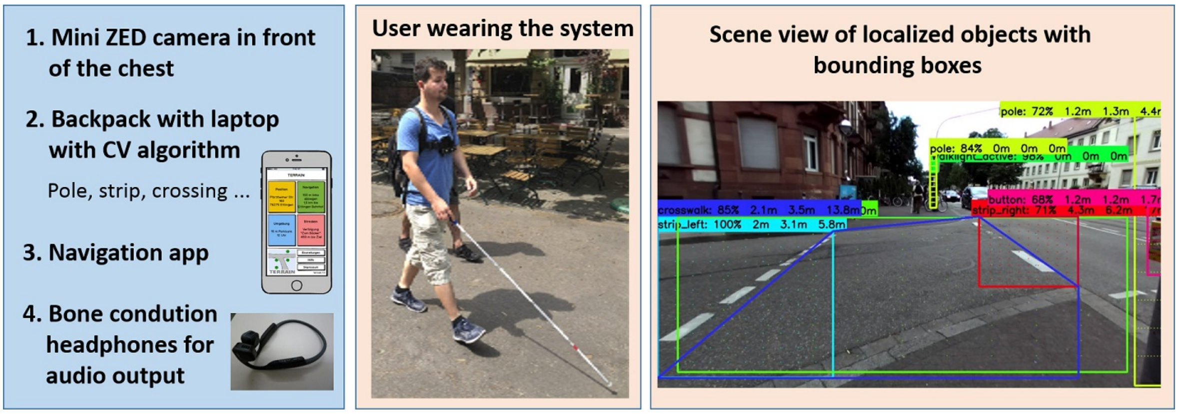


Figure 4.1.: The prototype consists of: (1) a stereo camera to capture images of the environment; (2) a backpack with a laptop on which the object localization algorithm was running; (3) a navigational iPhone app; (4) bone conduction headphones for audio output. In the middle of the figure, there is a picture (©Angela Constantinescu) of a participant wearing the system during the study and on the right-hand side one can see an image (©Daniel Koester) of a typical intersection crossing scene with localized objects, including their bounding boxes, existence probabilities and distances.

& Mobility training (Banovic et al. 2013), but unfamiliar areas as well as changes in the environment pose great challenges, affecting the mobility of people with visual impairments.

Our contribution is to enhance turn-by-turn navigation information with specific relevant details about crossings into one consistent interface, making a smooth transition from macro- to micro-navigation. To this end, together with the Terrain project partners, we developed a mobile prototype in a user centered design approach, involving a blind person and two mobility trainers during development. The prototype consists of (see 4.1):

1. A pedestrian navigation application relying on OpenStreetMap, designed specifically for people with visual impairments, taking into account roadside and crossings, implemented by iXpoint Informationssysteme GmbH and described in Ritterbusch and Kucharek (2018). The main aspects of their work is to find safer paths and crossings, giving more accurate path guidance on sidewalks, and offering additional information on street sides, crossings and their relevant accessibility features.

2. A computer vision algorithm that identifies and localizes zebra crossings and pedestrian traffic lights including push-buttons and signal state (red / green), implemented at KIT under the lead of Daniel Koester using deep learning approaches (Koester 2020). The computer vision algorithm was running on a Gigabyte P35 laptop with a GPU.

3. A single, **speech-based auditory display** based on the grammar described in subsection 4.1.1 and designed to handle both the navigational instructions and the crossing information from the computer vision system. In this section, we will focus on the auditory display only, but will shortly outline the other components as well, for a better understanding of the prototype used for evaluation.

Figure 4.2.: The hardware of the prototype: (1) iPhone running the prototype app; (2) stereo camera for capturing objects in the environment and their distance to the user, fixed on GoPro chest straps; (3) backpack with a laptop on which the computer vision algorithm was running; (4) bone conduction headphones for audio output, one pair for the participant and one for the test leader (image ©Angela Constantinescu, Daniel Koester)

In order to incorporate all this functionality coherently into one single system, the project partners from the iXpoint company[2] created a modular platform for distributed development on mobile devices, described in Ritterbusch et al. (2018). The auditory display was implemented in LUA 5.3, as one module of this platform. Further modules were necessary for: telling the user his current location; calculating the next actions to be conducted by the user given a route and the current position; computing POIs that are relevant to the user around a certain GPS coordinate; communicating with the crosswalk objects localization algorithm, which was a separate system running on a laptop. The implementation of all these modules was conducted by the working students Maximilian Awiszus and Jean Baumgarten under the supervision of Vanessa Petrausch and Angela Constantinescu.

The user actions implemented in this prototype were: *turn* (as in “turn right”), *follow* (when walking straight and no other events occur), *cross* (a street), *depart* (at route start), *arrive* (when reaching the destination), *abort* (for canceling the routing).

The hardware of the prototype consisted of (see 4.2): (1) an iPhone that was at the core, running the pedestrian navigation algorithm, connecting all other hardware components together, handling the user interaction and controlling the preferred settings chosen by the users; (2) a stereo camera strapped on the chest of the users that captured the environment and sent frames to the computer vision algorithm; (3) a laptop carried in a backpack that processed the captured frames, localized crosswalk objects in the frames, if any (traffic lights, push buttons, zebra crossings, etc.), and sent the information through Bluetooth to the mobile application running on the iPhone. The mobile application generated the speech instructions based on the grammar described in the previous subsection, 4.1.1, and transmitted these instructions via Bluetooth to the (4) bone conducting headphones AfterShokz Trekz Titanium (AfterShokz 2020) worn by the user.

Since the camera we used had a horizontal field of view of 90 degrees, when using the clock-face system, the direction of objects (zebra crossings, traffic light elements) mapped to 11, 12 or 1 o’clock, which was too inaccurate. Thus, we implemented a variation of the clock-face system to allow for more

---

[2] https://www.ixpoint.de/

accuracy by adding half hours in-between full hours. For instance, twelve thirty would correspond to 15 degrees right to the user (assuming that zero degrees is straight ahead), or better said, to an interval around this point, namely from 7.5 to 22.5 degrees. This idea is not new - it appears in the title of the German movie "Peas at 5:30" (Büchel 2004), which suggests that it might be used by orientation and mobility trainers in this country. According to this half-clock-face system, the horizontal angle of view maps to 10:30, 11:00, 11:30, 12:00, 12:30, 13:00 and 13:30. Each half-hour interval represents an angle of 15 degrees. In our case, having a field of view of 90 degrees, we only use half of the intervals for 10:30 and 13:30 (7.5 degrees on each side). The advantage of the half-clock-face system for conveying directions is that it is short and quick in German and English ("one thirty" would be "halb zwei" in German), more accurate than clock-face and left-right, and more intuitive than degrees.

Besides speech, we used sonification in form of brief sounds to inform the user when traffic lights, push-buttons or zebra crossings were at +/-3 degrees straight ahead. This was a quick way to inform the user when he was on a straight path towards an object and help him stay on this path. For that, we used a sound similar to the pilot tone of some traffic lights in Germany.

The following example describes the compact output of our system combining navigational speech instructions (NAV-Speech) with information about objects in the environment obtained from the computer vision module, conveyed both as speech (OBJ-Speech) and as sonification (OBJ-Sonification). This makes the smooth transition between components (navigation, objects localization) and also between macro- and micro-navigation assistance more obvious. The English translation is provided below (the translation shows the different word order of the two languages):

German (original):

> [NAV-Speech]: 47 Meter folgen.
> [NAV-Speech]: In 30 Metern Hauptstraße an Zebra überqueren.
> [NAV-Speech]: In 11 Metern überqueren.
> [OBJ-Speech]: Zebra auf halb zwölf in 8 Metern.
> [OBJ-Speech]: Zebra auf zwölf in 6 Metern.
> [OBJ-Sonification]: tac-tac-tac-tac *(sound plays shortly after previous instruction starts being spoken, so Speech and Sonification can be heard simultaneously)*
> [OBJ-Speech]: Zebra auf zwölf in 3 Metern.

English (translation):

> [NAV-Speech]: Proceed 47 meters.
> [NAV-Speech]: In 30 meters, cross Main Street at the crosswalk.
> [NAV-Speech]: Cross in 11 meters.
> [OBJ-Speech]: Crosswalk at eleven thirty in 8 meters.
> [OBJ-Speech]: Crosswalk at twelve in 6 meters.
> [OBJ-Sonification]: tac-tac-tac-tac *{sound plays shortly after previous instruction starts being spoken, so Speech and Sonification can be heard simultaneously}*
> [OBJ-Speech]: Crosswalk at twelve in 3 meters.

The information about crossings is thus obtained from two different sources: (1) the navigation instructions contain information about the presence of a crossing, obtained from the underlying map. The distances are only approximates, given localization accuracy. E.g.: “In 30 meters, cross Main Street **at the crosswalk**”; (2) the computer vision algorithm then identifies the object (crosswalk) in the images captured with the video camera and gives precise information about its position and distance: “**Crosswalk** at eleven thirty in 8 meters.”

4.1.3. Evaluation

We conducted a qualitative user study to learn whether the additional information about crosswalks supports users in difficult crossing situations, without cognitively overwhelming them: were the participants more accurate in finding crosswalks using our prototype in addition to their long rehearsed, own methods? Is the extra information about crosswalks useful? How high is the cognitive load during navigation with the additional information? Additionally, we wanted to uncover possible problems with the grammar and to get suggestions for improvements. In order to answer these questions, we tested our system with people with visual impairments.

We first conducted a pre-evaluation with two mobility trainers and two accessibility experts. An additional accessibility expert who is also blind acted as participant. As a result, the initial routes were slightly changed. Moreover, guidelines for the sighted escorts were established and the test procedure was adjusted. Sighted escorts are sighted people who accompanied each participant and provided for their safety during the test.

Fifteen people with visual impairments took part in our study - we will refer to them as P1 through P15. Four of them were female and eleven male, aged 21-71, with an average age of 43 (SD 17.5). Four were congenitally blind, three late blind and eight severely visually impaired. Eight said they could see neither traffic lights, nor zebra crossings (or too little, not by daylight). Six participants could not discern traffic light signals, but could see zebra crossings, mainly conditionally (very weak or only within a small distance). Only one participant (P1) said he could see both, although traffic light signals only occasionally. All fifteen said they mainly travel by foot, and eleven said they additionally use public transportation. One walks (almost) always alone, thirteen walk alone often, and one walks only sometimes alone. Thirteen use navigation aids, though some of them conditionally (only in the city, in unknown areas, when they search for something). Thirteen used a white cane during the study, while two did not use any (one did not need a white cane, the other one intended to buy one soon). Table 4.1 summarizes the general information about the participants. The ages and onset of the visual impairment are specified as intervals for reasons of anonymization: since the community of people with visual impairments is rather small, the identity of a person could be inferred when revealing the exact age and onset of the visual impairment next to all other demographic data.

We asked the participants to walk two fixed routes through a busy urban area, containing several traffic lights and zebra crossings (see Fig. 4.3). A third route was used in addition, prior to the test, for training and familiarization with the system. Most of the participants (12) did not know the area in which the tests took place, and three knew very little about it. According to their own statements,

ID	Age Span	Gender	Visual Impairment	Onset (Age Span)	Mobility Aid
P1	40-49	m	blind[a]	20-26	cane
P2	40-49	f	low vision	20-26	cane
P3	70-79	m	blind	birth	cane
P4	20-29	m	blind	birth	cane
P5	60-69	m	low vision	birth	*none*
P6	30-39	m	low vision	20-26	cane
P7	20-29	m	low vision	birth	cane
P8	50-59	m	low vision	20-26	*none*
P9	30-39	m	low vision	12-15	cane
P10	70-79	f	low vision	20-26	cane
P11	30-39	f	blind	12-15	cane
P12	20-29	m	low vision	birth	cane
P13	40-49	f	blind	birth	cane
P14	40-49	m	blind	birth	cane
P15	20-29	m	blind	12-15	cane

Table 4.1.: Demographic data of the participants of our user study: the participants' ID, age, gender, type of visual impairment, onset of the visual impairment, and the use of a white cane

---

[a] Blindness is defined differently in various countries. The definition of blindness in Germany is identical with the proposed WHO category 4 (WHO 2020)

none of the participants recognized the actual routes taken during the study. These are the routes (see Fig. 4.3):

- **routeTraining** (770 Meters): was chosen to be as short as possible while containing at least one zebra crossing and one traffic light, which could be crossed multiple times for training purposes.

Figure 4.3.: The two test routes walked during the user study. "RouteForth" goes from the upper right corner to the lower left, and "routeBack" the other way around. The underlying map was downloaded from https://saferoute.ixpoint.de/pedestrian-routing.

- **routeForth** (580 meters): the first actual test route, or “route forth”.
- **routeBack** (580 meters): the second test route was the same as the first route but in the opposite direction.

RouteForth and routeBack were chosen to: (i) be equivalentto each other; (ii) be as short as possible; (iii) contain at least two zebra crossings and two traffic lights; (iv) contain little dangerous, unsupervised crossings; (v) be a real scenario; (vi) contain a quiet place in-between where one could sit down and fill in questionnaires; (vii) be close to our offices or at least to a tram station, to reduce travelling times to and from the test for everyone involved.

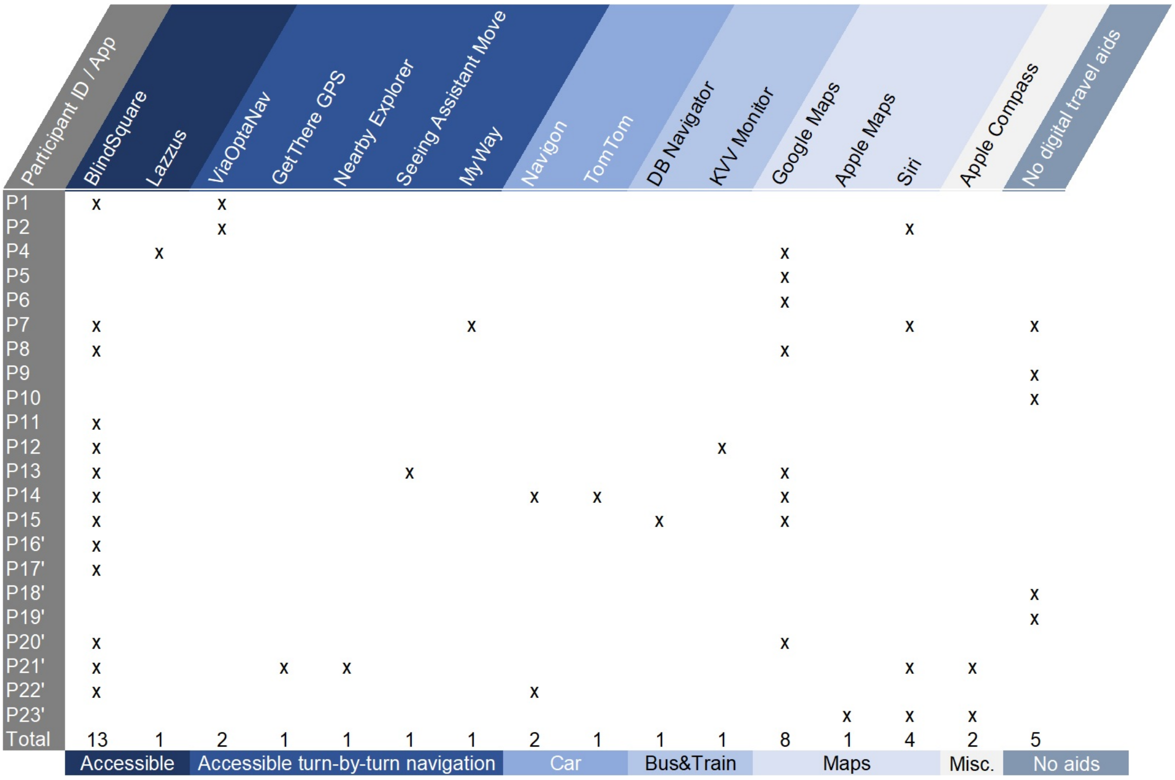
For a reproducible test setup and for covering as many traffic lights and zebra crossings as possible on a short route, the paths used during the user study were fixed GNSS trails hard-coded in the app, consisting of several routing segments. These routing segments followed the rules of the developed routing solution, i.e. they simulated the real routing. The rules to generate the speech instructions were used afterwards just as for any other route calculated by the system.

On each of the two test routes, the participants walked either with the navigation information alone (**routeNav**) or with the navigation *and* the additional information about crosswalk objects (**routeNavObj**), in interchangeable order. Seven of the fifteen participants walked forth without the additional object information and back with it and eight the other way around. There was no apparent learning effect, as shown by an analysis of the results after the study.

Before the test, the participants signed a declaration of consent and filled in a user background questionnaire. During and after the test, further questionnaires were filled in, as well as a confirmation of payment at the end of the test (each participant received a compensation of 30 Euro).

During the test, a sighted escort was near the participant at all times and provided for his or her physical security. For the fifteen tests, three times the escort was a mobility trainer, four times it was someone who had experience in escorting people with a visual impairment, and eight times it was someone with little or no experience in escorting at all. All escorts received and read a document written by the mobility trainers, which contained rules on how to behave during the test, as well as another document on how to escort a person with visual impairments. We also agreed on a procedure that had to be followed before crossing the street, which implied that the participants stop when a traffic light or zebra crossing was announced, and wait until the test leader made some notes and controlled the prototype (manually started the computer vision algorithm to identify the object announced by the navigation module, i.e. traffic light, push-button or zebra crossing). Then the users had to position themselves for the crossing, which meant finding the beginning of the crossing and identifying the direction to cross. For security reasons, the users had to wait until the test leader gave the OK for crossing, even after identifying a good timing (e.g. walk phase) and only cross in the company of another person. This procedure was explained to the participants before the test.

The speech instructions were given using the iPhone’s Siri voice. In the beginning (the first two participants), the feedback was given through two bone conduction headphones, one worn by the participants and another one by the test leader, using a Bluetooth splitter connected to the iPhone. However, during these tests, the connection to the participants’ headphones was often lost, and



Participant ID / App	BlindSquare	Lazzus	ViaOptaNav	GetThere GPS	Nearby Explorer	Seeing Assistant Move	MyWay	Navigon	TomTom	DB Navigator	KVV Monitor	Google Maps	Apple Maps	Siri	Apple Compass	No digital travel aids
P1	x		x													
P2			x											x		
P4		x										x				
P5												x				
P6												x				
P7	x						x							x		x
P8	x											x				
P9																x
P10																x
P11	x															
P12	x										x					
P13	x					x						x				
P14	x							x	x			x				
P15	x									x		x				
P16'	x															
P17'	x															
P18'																x
P19'																x
P20'	x											x				
P21'	x			x	x									x	x	
P22'	x							x								
P23'													x	x	x	
Total	13	1	2	1	1	1	1	2	1	1	1	8	1	4	2	5
	Accessible		Accessible turn-by-turn navigation					Car		Bus&Train		Maps			Misc.	No aids

Figure 4.4.: Use of navigation apps by 22 people with visual impairments. P1-P15 are the participants to our user study, same as in Table 4.1. P16'-P23' are participants in the other two user studies.

reestablishing was troublesome, so that we decided to use a speaker for the rest of the study (P3-P15). The iPhone and the speaker were held by the test leader, who controlled the prototype. The participants only heard the instructions from the speaker and had to decide how and where to walk. When walking on routeNavObj, the participants additionally carried a backpack containing a laptop (Fig. 4.2) on which the objects localization algorithm was running. The backpack's chest straps had the ZED mini camera attached. On routeNav the backpack was not needed and was carried by the test leader.

During the test, we gathered the following data from each user: (i) *Questionnaires*: a user background questionnaire, two Raw NASA-TLX (Hart 2006), two System Usability Scales (Brooke 2013), and general questions about the prototype; (ii) *Log files*: log files recorded by the App, containing among others the speech instructions that were spoken, GPS tracks and the times to complete each route; log files recorded by the laptop containing output of the computer vision algorithm and status of the Bluetooth connection; (iii) *Videos*: when the computer vision algorithm was on, the laptop recorded videos taken from the participant's camera at crossings; (iv) *Comments*: the test leader noted all relevant comments made by the participants during the test.

Usage of mobile applications. In our questionnaire, we asked the participants what kind of tools they use to support them during travel. Fourteen out of 15 participants answered the question. To these, we added 8 answers to the same question from two other user studies conducted at our center, giving a total of 22 answers. The aids used can be divided into five categories: (1) accessible Point of Interest (POI) based navigation, (2) accessible turn-by-turn navigation, (3) car navigation tools, (4) public transportation and (5) maps. The results are presented in Figure 4.4. For accessible POI based navigation, BlindSquare and Lazzus were used. Some participants used the accessible turn-by-turn navigation apps ViaOptaNav, Seeing Eye and SeeingAssistant Move. Others also used the car navigation apps MyWay and Navigon and for public transportation the apps DB Navigator, an app that provides information about public transport in Germany, and KVV, a local app of the tram network in the city of Karlsruhe. Moreover, the participants used the apps GoogleMaps and AppleMaps. One participant (P3) said he does not use his phone's apps (he had an iPhone) because of overwhelm, too much information. The analysis shows that the most used application was BlindSquare (13 people), pointing out to the importance of objects or points of interest for the orientation of people with visual impairments. GoogleMaps was the second most used app (8 people), although it is not specifically designed with accessibility in mind and it does not include all the features that people with visual impairments need. Accessible navigation apps were used by 6 people, but almost each one used another app. There were 15 apps in total used by 22 people with visual impairments, and each of these apps comes with its own interface. A standardized interface would reduce the learning curve and would encourage more users to use digital aids during travel. Moreover, each participant named more than one app, most named 2 to 3 apps that they actively used. This suggests that users need more than the functionality provided by one app alone. Combining and integrating smoothly more functionality into one single app could also increase the number of app users, which, in turn, increases the mobility of the target population. Nine out of 22 users, for instance, use a combination of POI and navigation apps. This suggests once again the importance of getting both POI and navigation information, which we considered in our grammar.

Orientation during crossing. We counted how successful a person was in finding a crossing with and without the additional feedback about objects at crossings. We consider it a "success" if the person aligned correctly for crossing. For each road that had to be crossed at a traffic light or zebra crossing, the number of times a person found the crossing was counted as follows:

- 0, when the participant did not manage to find the (correct) crossing or when a hint was given.
- 1, when the participant successfully found the crossing without hints.

The total number of crossings found on routeNav is 54 out of 60 (mean 3.6 per participant) and on routeNavObj 55.5 out of 60 (3.7 per participant). The difference is not statistically significant, p=0.56 in a paired t-test. While this result was somewhat disappointing, it was also to be expected, given the short training time available. Sánchez and de la Torre (2010) found in a user study of a navigational aid device for people with visual impairments that "practice and increased dominion over the system was key to coming to trust the software in order to navigate their way to the destination." This is in line with our observations during our user study. The test leader noticed that in very complex situations such as high traffic, crowds or atypical intersections, some users did not react to the instructions given by the prototype, and apparently tried to solve the problem using their own, well-rehearsed methods, but did not always succeed. For the same reason, we did not compare the data gathered about the time taken to complete a route. Studies such as that of Zeng et al. (2012) report an increase in time while using a new system when training time is short. However, Zeng et al. (2012) also report a considerably improved safety.

If finding the crossing did not seem to be (statistically) better with the additional information, when crossing the street the test leader could observe an improvement. In at least two cases (P7, P15), it could be observed how the participants oriented themselves before and during zebra crossings better when using the CV system than without it. They started more aligned with the crosswalk and their path over the street was more straight when using the CV system, according to the test leader's observations, confirmed by the participants' own comments. Moreover, one could occasionally see how the system helped users find the crossings easier, without going too close to the street edge to orient themselves. This was the case with at least two participants, who walked too close to the traffic lane on routeNav. The situation was classified as dangerous by the test leader because the cars were coming from behind at high speed and the lane was very narrow. Moreover, even the three participants who could see both zebra crossings and traffic lights pole (but could not see the traffic signal) to an extent where they could orient themselves without any help, still liked to receive information from the CV system. First of all as a means of reassurance, but also in order to assess the traffic lights state at inaccessible traffic lights or when they could not find the (vibrotactile) push button.

Usefulness. When asked whether they preferred walking with the additional information at crossings or without it, all fifteen participants answered they preferred walking with it. One of them (P13) even mentioned that she felt more secure while crossing a street with the additional information than without it. This indicates that the system provided important additional information that was not otherwise available to users. Also, both the computer vision system and the navigation App were found to be useful with a mean score of 1.7 for CV system and 1.9 for the navigation App on a scale from 1 (very useful) to 5 (not at all useful). Three participants said that in their answers they did not take into account the heavy hardware (laptop) and the positioning inaccuracy of the navigation. Two other participants said they would have given a better score if the conditions mentioned above (less hardware and better positioning) were fulfilled.

Usability. The system's usability scale (SUS) score for routeNav is 69, which is just around average for mature systems (Bangor et al. 2008). The computer vision system received a smaller average SUS score of 63.5 with p=0.051 in a paired t-test, which can be explained by the heavy hardware (the full backpack weighed about 12 pounds) and the partial instability of the Bluetooth connection during the tests. Computing the Pearson correlation coefficient between the SUS scores differences (routeNav minus routeNavObj) and number of Bluetooth crashes showed some correlation ($\rho = 0.42$). Given that our system is in a prototype phase and not a mature system, we consider this result to be satisfactory.

Cognitive load. The participants completed a raw NASA-TLX questionnaire after each test route (only the rating scale without the weights). The average Raw TLX values for all 15 participants were slightly higher for the routes walked with the navigation App and the computer vision system (routeNavObj) than for the routes walked with the navigation App alone (routeNav), namely 4.9 out of 20.0 (24.4%) for routeNavObj and 4.7 (23.7%) for routeNav. The difference, however, is not statistically significant (p=0.66) according to a paired t-test. There was also no statistically significant difference between the Raw TLX average values for routeF - 4.7 out of 20.0 (23.2%), and routeB - 5.0 (24.7%), p=0.423. In NASA-TLX benchmarks, sighted people have for navigation tasks (planning and following a route) a minimum cognitive load of 19.72 and a maximum of 68.90, and raw (RTLX) scores are slightly higher than the weighted ones (Grier 2015). Thus, the cognitive loads of our users were very low. It is possible, however, that the test conditions, including the presence of the test leader and sighted escorts made participants feel safer and more relaxed, as they could not get lost or miss an appointment like in a real-life scenario. All this could have contributed to a lower cognitive load. Another bias might come from the background of the participants: all except for one walk often alone, and only one from 15 walks *sometimes* alone. This suggests that at least 14 from 15 participants are highly mobile and independent when travelling. One way to avoid such biases in the future is to mention in the study invitation that people with reduced mobility are explicitly asked to participate.

Users' feedback. We collected the users' feedback from comments made spontaneously by the participants during the test, as well as from the questions asked after the study: "What did you like/dislike most of the entire system?". We conducted a thematic analysis (Braun and Clarke 2006) and found the following themes:

- *Pedestrian navigation.* The pedestrian navigation was appreciated, provided that the localization works. Two participants liked the fact that the street sides and street names were mentioned.
- *Additional objects information at crossings.* Ten out of fifteen participants found the additional information provided at crossings, or the idea, or parts of it, to be good: "Without the instructions, I might not have found the zebra crossing" (P9), "I loved the way it noticed traffic lights and zebra crossings! Especially at zebra crossings, one can orient oneself well" (P7), "I felt more insecure without the [additional information]" (P13), "The detection of zebras worked very well, I liked it very much" (P15), "[I liked] that it announced traffic light state, especially where there's no push button" (P2), [I liked:] "the traffic light state recognition" (P8), "the traffic light state announcement" (P14), "...when it said: Zebra, push button, traffic light" (P12), "the possibility to recognize zebras and traffic lights", etc.
- *Dislikes.* There were several things that the participants disliked: besides the already mentioned positioning inaccuracies and heavy hardware, the bone conducting earphones were a problem for one participant, who said they did not work so well with his hearing aid.
- *Speech output.* Five participants said that the speech instructions were good (short, concise, clear, have few error sources), and others commented positively on parts of the output: the compact grammar, the repeating "follow" instructions every 25m, the fact that the feedback was consistent for the two systems, the directions for both systems using "slightly/-/sharp left/right" and the half-clock-face system. The participants also made proposals on how to improve the user output as well as the system in general. Thus, participants wished the system could also identify: obstacles in general (P3), lampposts (P6) and poles (P8, P9), puddles, dog excrement and placards (P10), flat curbs (P13) and construction sites (test leader). The participants also wished that important instructions (such as those announcing crossings or turns) have a higher priority and be able to stop instructions that are not so important. This is because during the test, it sometimes happened that the information about bad GPS signal, which was very verbose, started right before or during the crossing, which sent instructions from the CV system into a queue that got played sequentially and too late.
- *User interface.* Some of the participants suggested various improvements of the user interface:

- Announce in advance turns that come after a crossing, or simply say before the crossing how it goes on afterwards (P5, P7, P15)
- Use round numbers for the approximate distances in the navigation instructions: 5m, 10m, 15m, etc (P2); or less specific terms: “soon”, “shortly”, “at the next opportunity”, “after the next intersection/junction”, “the destination is nearby” (P1, P13)
- Have the option to take out street names from the instructions (P1, P13)
- Reduce the number and verbosity of the instructions about positioning inaccuracy (P5, P13)
- Have a way to repeat instructions (P5)
- For fast walking users[3], there should be an option to speed up the speech and /or an alternative interface based on speech *and sounds* that is faster than speech alone (P4). The participant suggested for objects identification to announce the object only once through speech: „Crosswalk in 8 meters at 12 o’clock“, and afterwards only via sonification, using a sound that gets increasingly louder or faster.
- Show traffic lights in the vicinity as points of interest (P8)
- Say the next crossroads, like in BlindSquare (P1)
- Integrate RTB‘s LOC.id[4] in our App (P4). RTB‘s LOC.id offers blind users a way to query traffic lights state and leads them over the street, but needs additional hardware installed at the traffic lights.

Discussion. We identified several issues that need to be considered in the research and development of future similar solutions. We discuss here our observations.

Positioning accuracy. The GPS positioning accuracy was quite bad throughout the test. An analysis of the speech logs from the two test routes, routeForth and routeBack, showed that 27% of all navigational instructions were omitted because of an inaccurate location signal (in these cases, the test leader would speak out the instructions during the test). Our positioning solution was however a state-of-the-art implementation for mobile devices, and we do not expect other systems to perform much better in similar conditions. One of the reasons why the positioning failed was the presence of high rows of buildings along the route - which is a common situation in urban areas. According to van Diggelen and Enge (2015) the mean GPS accuracy *in the city* is about 16-17 meters and can be directly correlated to the height of the surrounding buildings. An improvement of the positioning accuracy is thus crucial for the future development of similar systems for this target user group. At the moment the study

---

[3] One participant, who was a young, born blind person, was walking very fast. So fast, that the sighted test attendants could hardly keep up with him. He could sense the presence of a very thin pole from more than one meter away, although he could not see it. He was very mobile and often walked unattended, exploring unknown environments. According to the mobility trainers, such a case is not unique. There are other known cases of totally blind people who even ride bicycles independently (Kish, Daniel 2011).

[4] https://www.rtb-bl.de/RTB/en/blind-aids/loc-id/technologyapplication/

was performed (2018), there was no low cost solution available for precise positioning, as needed by people with visual impairments.

Orientation of camera. The straps on which the camera was attached on the chest also posed problems, mainly for women, but also for some men. In these cases, the camera was oriented too high and did not “see” zebra crossings, so it had to be held into the right position with one hand or by using something that was at hand, e.g. a small package of tissues or a wallet. An adjustable holder would make the camera stand out even more, and poses social acceptability problems. An alternative could be to use a camera integrated into sunglasses or the smartphone’s camera, in which case the phone should be worn in front of the chest, attached with an adjustable neck strap.

Situational awareness. Participants reported that the system provided information (e.g. about turns or localization accuracy) right in the middle of a crossing, which is unacceptable. People with visual impairments must concentrate on crossing safely and should not hear other, less important instructions while crossing. The two systems should be better synchronised to support the situational awareness: if the CV is switched on and detects an object within a certain range (e.g. the beginning of a crosswalk or a push button within 5 m), navigation should suspend other instructions.

Alternative interface. While most users appreciated the interface, the test leader noticed that the system was apparently slowing one participant down, while another participant (P4) was annoyed by the slow feedback and usually moved much too fast for the pace of the interface and even for the pace of the sighted experimenter. We thus argue that an alternative, faster interface is essential for the adoption of the system by fast users.

Test procedure. During the tests, we had planned to gather more objective data (such as the time needed to cross the street, distance/steps walked by the participants), but it turned out to be impossible due to safety issues or last-minute adjustments. It is therefore essential to film the entire study, use a pedometer, log the session and record the comments for later transcription in order to collect more objective data.

During our tests mobility trainers, accessibility experts and people with little experience in guiding blind persons were involved to keep the participants safe. We have observed that even accessibility experts with years of experience with people with visual impairments react differently than mobility trainers in difficult and confusing traffic situations. They intervene more often and thus influence the outcome of the study. Thus, we have learned that, in safety relevant situations of navigation, it is essential to involve mobility trainers not only in the study design, but also in each individual test involving crossings or other dangerous situations, which is in line with the findings of Williams et al. (2014).

Users change their mind. One participant (P5) said during the study that he does not need "follow" instructions. However, at some point during the study he asked: “how much longer?”. Later on, he said he found the "follow" instructions reassuring. What we learned out of this is that users sometimes don’t know what they want. Thus, instead of asking users how a future system should be, researchers should observe what users do, prototype (starting with cheap, low level prototypes) and repeatedly evaluate. This is consistent with Nielsen, Jakob (2001), who encouraged researchers to “pay attention

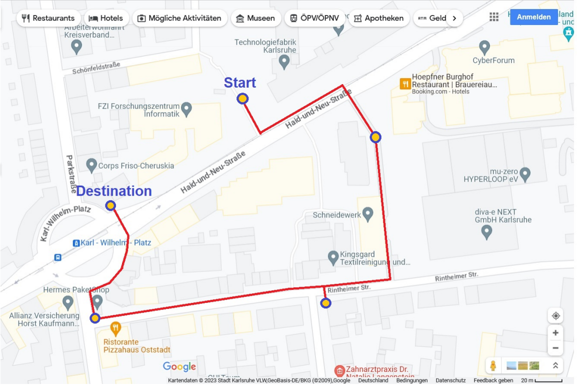


Figure 4.5.: The route walked with Google Maps and with our app, which implements parts of our grammar. The underlying map is downloaded from `https://www.google.de/maps.`

to what users do, not what they say. Self-reported claims are unreliable, as are user speculations about future behavior. Users do not know what they want."

4.1.4. Comparison with State-of-the-Art

Among the turn-by-turn navigation apps, Google Maps is the one most used by people with visual impairments (see Table 4.4). We thus decided to compare the output of our prototype, which implements parts of our grammar, with the output of Google Maps. In June 2019 we conducted a small test, in which we walked one fixed route, about 650 meters long, through a city, once with our prototype and once with Google Maps (with the pedestrian option). We chose the route to include several crossings. Thus, the route is artificial, including crossing one street forth and immediately back - see Figure 4.5. We thus had to hardcode the route in our app. For Google Maps we defined intermediate points to make sure that the routes coincide.

In Table 4.2 we compare the output of the two apps. The output of our app was updated according to the results of the user study described in the previous Subsection, 4.1.3, in that distances were rounded to nearest 5. Additionally, some "follow" instructions, which appeared right before or right after another instruction such as "turn" or "cross", were removed from the output. This is because "follow" instructions should only occur when no other instructions follow shortly. The reason why they appeared is possibly due to the positioning data jumping around. In Google Maps, we also added a couple of instructions, which were not spoken due to missing GPS signal. At times, we experienced the same problems with our app, mostly in the same areas. This suggests that standard positioning signal and accuracy is currently still too low in cities for proper use in applications for people with visual impairments. The output of both apps was originally in German and we translated it here to English for exemplification only. Instructions in Google Maps about the intermediate stops, which are irrelevant to the comparison, are listed in gray.

One can see several differences between the instructions:

- Our grammar includes "cross" instructions and mentions the presence of crosswalks (such as zebra crossings) or traffic lights with their accessibility features such as acoustic signal. Google Maps mentions the street side on which the destination is and leaves crossing streets up to the

Event	Our grammar	Google Maps
Start	Computing route: 630m. Start at 9 o'clock on footpath, left side, towards Haid-und-Neu-Straße. [The phone vibrates 1 second long when the user reached the walking direction]	Head south-east toward Haid-und-Neu-Straße
Follow	Follow 20 meters	
Turn	In 15 meters turn left onto Haid-und-Neu-Straße	Turn left onto Haid-und-Neu-Straße
Turn, cross	In 30 meters cross right Haid-und-Neu-Straße	Turn right
Follow	Follow 100 meters	
	Follow 80 meters	
Interm. dest.		[3-sounds chime] You have reached your destination
Follow	Follow 50 meters	
Turn	In 30 meters turn right on Rintheimer Straße	Turn right onto Rintheimer Straße
	In 10 meters turn right	
Turn, cross	In 35 meters cross left on Rintheimer Straße at crosswalk	
Turn, cross	In 15 meters cross left	
Interm. dest.		[3-sounds chime] The destination is on the left side
Turn	In 5 meters turn right on Rintheimer Straße	
Turn, cross	Now cross right on Rintheimer Straße at crosswalk	
Follow	Follow 170 meters	
	Follow 140 meters	
	Follow 120 meters	
	Follow 90 meters	
	Follow 65 meters	
	Follow 40 meters	
Turn	In 30 meters turn right on Georg-Friedrich-Straße	Turn right onto Georg-Friedrich-Straße
	In 15 meters turn right	
Interm. dest.		[3-sounds chime] You have reached your destination
Turn	In 10 meters turn right on Karl-Wilhelm-Platz	Turn right onto Karl-Wilhelm-Platz
Follow	Follow 35 meters	
Cross	In 30 meters cross Haid-und-Neu-Straße at traffic light	
Destination	In 15 meters destination reached In 10 meters destination reached	[3-sounds chime] The destination is on the left side

Table 4.2.: Comparison between the pedestrian navigation instructions generated with our grammar and Google Maps, which is at the moment state-of-the-art, meaning most used app by the visually impaired community.

user, which for sighted users is perfectly fine. For visually impaired users, however, this is not enough, especially in unknown environments.

- Upon start, our grammar tells the user in which direction to turn in an egocentric way: “9 o’clock” means the user must turn 90° to his left. Google Maps gives the allocentric direction: “south-east”, which is less error prone, but it assumes again that the user *sees* on the map which direction that is. Our grammar also mentions the side of the road on which one should start. This is to reassure the user that he starts in the correct direction and on the correct street.
- Our grammar includes “follow” instructions, which, as we have seen in the evaluation, are found reassuring by some users. They also inform users that the system is still up and running.
- Our grammar mentions the distance left until an event: “In 30 meters turn right”. Google Maps probably assumes that one sees the map and the distances on the screen, which is not necessarily true for people with visual impairments.
- Our grammar repeats “turn” instructions, in a shorter version (without the street name), to make sure that the user does not miss it and also to account for possible localization fluctuations. This is another feature that is very important for visually impaired people, since missing a turn can be very frustrating, if not dangerous.
- To compensate for the amount of extra information, our grammar combines turn and cross instructions when these occur concomitantly: “cross right”. In German, the direction comes first, which follows the order in which the actions are performed and relieves users’ cognition.

Thus, our grammar, although more verbose than the state-of-the-art (Google Maps), provides the visually impaired users with information that is paramount for the orientation on the street. Moreover, Google Maps requires that the user looks occasionally at the screen, which, in high traffic areas, can pose a safety risk even for sighted users. We thus argue that navigation instructions based on our grammar are better for people with visual impairments than the state-of-the-art. Additionally, our grammar could be useful to sighted people as well, by allowing them to navigate more safely without looking at the smartphone’s screen.

4.1.5. Conclusion

Based on a literature review and with the involvement of 5 experts, we defined a grammar for conveying navigation instructions enhanced with environmental and accessibility information for people with visual impairments during *outdoor* travel. In a user-centered development approach, with the involvement of a blind person and two mobility trainers throughout the design process, we integrated our grammar into a prototype created to support people with visual impairments during outdoor travel. The prototype provides users with: (1) navigation guidance tailored to their needs, but with approximate distance estimates in the city; (2) additional information about crossing elements including their *precise* position and distance to the user; and (3) a single, tailored speech interface based on our grammar, with instructions that are short, concise and naturally sounding, while still conveying all important information. In a user study with 15 people with visual impairments we found that the idea of the entire system, and especially the precise identification of crossing objects was very much appreciated. One of the participants even stated that she *felt safer* with the system than without it. The prototype helped some of the participants to find the crossing without going too close to the traffic lane, and to better keep aligned before and during crossing. All participants preferred walking with the additional information and found both the navigational instructions and the additional information useful. This suggests that users wish to have more information and more precise than is currently available in navigation solutions. The additional information came at a low cognitive cost, as the subjective cognitive load was only slightly higher when walking with the additional information at crosswalks. This could also be due to the consistent interface, which was appreciated by users. Finally, users made suggestions on how to improve the interface. We believe that the results of our study will be useful for the future development of improved speech-based interfaces for navigation and crossing assistance for people with visual impairments.

4.2. Speech Auditory Display for Indoor Travel: Grammar for POI-based Descriptions of the Surroundings

This section is based on our ICCHP publication Constantinescu et al. (2022b) © Springer and the master thesis of Eva-Maria Neumann (Neumann 2022), which was jointly supervised by Angela Constantinescu and Dr. Karin Müller.

In this section we address the question of *how* to convey information to people with visual impairments during *indoor* travel. More precisely, we address especially blind people in this section. While people with low vision could also profit from the extra information provided, in limited indoor spaces they have fewer problems when navigating than blind people. Thus, we focus our attention only on blind users and evaluate our approach accordingly.

When travelling indoors, blind people mainly rely on a sighted assistant or on asking sighted people on site. However, such instructions are seldom clear, as “sighted people often do not understand what kind of instructions are most helpful” (Gomez et al. 2016) to people with visual impairments. Moreover, blind employees from Chile admitted in a user study that “it took them a long time to get to know the workplace and some of the participants even disclosed that they only knew how to get to the essential areas such as the workstation, kitchen, toilet, lifts, and the way out” (Gomez et al. 2016) - thus, they hardly knew a building that they visited almost daily. Knowing the environment allows people to make better and more spontaneous decisions and improves their quality of life. Engel et al. (2020) showed that people with blindness would like to be more independent when travelling indoors and would like to use textual descriptions of buildings if they were available. While several navigation systems for indoors do exist, they do not achieve large-scale deployment due to cost, accuracy and usability, including the lack of a robust feedback modality (Fallah et al. 2013). Finally, Striegl et al. (2023) showed that users form better cognitive maps when using indoor orientation systems based on landmarks than with classic turn-by-turn navigation systems. Yang et al. (2011) also found that providing environmental information improves spatial awareness of people with visual impairments and thus, fosters their independence. Thus, we decided to also focus on conveying orientation information based on accessibility features (objects) in the environment and not on turn-by-turn navigation.

In this section we analyze how to automatically generate egocentric descriptions of objects in the environment from OpenStreetMap indoor maps to help blind users orient themselves in unknown buildings. We create and evaluate a German grammar in a user-centered way, making sure that it fulfills the strict requirements of the users.

4.2.1. Analysis of Literature and Existing Solutions

Navigation systems have been in the focus of research for over 30 years now. There are several surveys that give a good overview, from various points of view. In 2020 alone, there were multiple journal articles that reviewed navigation solutions for people with visual impairments either indoors (Plikynas et al. 2020, Kandalan and Namuduri 2020, Simões et al. 2020), or both indoors and outdoors (Kuriakose et al. 2020). A couple of older articles (Fallah et al. 2013, Huang and Gartner 2009) review systems for the general users, but solutions for people with visual impairments are also included. These complement very well the newer reviews, which sometimes look at the previous five years only (Kuriakose et al. 2020, Plikynas et al. 2020). However, the interface is hardly ever in the focus of attention and sometimes not at all considered. In Nicolau et al. (2009b), a grammar for *turn-by-turn instructions* is created in a user centered approach by analyzing route descriptions made by blind people. Objects on the route are mentioned, but the focus is on navigation, rather than a description of the immediate environment. In Madugalla et al. (2020), textual descriptions of floor plans are automatically generated, in an allocentric approach. This means that the entire plan is described independently of the user perspective. A very similar approach had been previously tried by Paladugu et al. (2012) as well. While such descriptions are useful for *planning* a trip to an unfamiliar building, they are not useful for on-site orientation, where egocentric information is needed. The approach that comes closest to specifying a grammar for egocentric textual descriptions of indoor spaces, like we do, is the one by Yang et al. (2011). Their Talking Points 3 system reads aloud a list of points of interest (POIs) around the user. Their system can be used both indoors and outdoors. Although they do not formally specify a grammar, they describe many of its components. Some details, however, remain unspecified. Our work complements their efforts and tries to close this gap. We include newer findings with respect to parts of the grammar and answer open questions in a user-centered way. Moreover, we focus on a solution tailored to the German language.

For specifying our grammar, we first performed an analysis of previous literature, including the reports of one online user study and a focus group previously carried out within our project. Additionally, we studied the output of three existing apps for indoor navigation and orientation: Blindsquare (MIP 2020), BFW Smartinfo (Berufsförderungswerk Würzburg), RightHear (RightHear). Our goal was to analyze the existent grammars and best practices and deduce a grammar for conveying indoor information to people with visual impairments. Blindsquare is a paid iPhone/iPad app that lists points of interest in the surroundings. It was specifically designed for people with visual impairments, and it is the most used app by the community for this purpose, as can also be seen in Figure 4.4. Unfortunately, there is no Android version available. Blindsquare was initially designed to work outdoors, but recently, it

Textual description rule	**Reference**
Object then Direction	Blindsquare (**POI**,**Ger**,**in**/out), Constantinescu et al. (2019) (**POI**/nav,**Ger**,out)
Direction then Object	Yang et al. (2011) (**POI**,**in**), BFW SmartInfo (**POI**,**Ger**,**in**/out), Blindsquare (**POI**,**Ger**,out)
Distance then Direction	Constantinescu et al. (2019) (nav,**Ger**,out), Blindsquare (**POI**,**Ger**,**in**/out)
Direction then Distance	Nicolau et al. (2009b) (nav,**in**), Constantinescu et al. (2019) (**POI**,**Ger**,out), Jain (2014b) (nav/**POI**,**in**)
Object should contain details	May et al. (2020) (**POI**,**in**)
Distance in fixed unit (e.g. meter, feet)	Yang et al. (2011) (**POI**,**in**), Nicolau et al. (2009b) (nav,**in**), Blindsquare (**POI**,**Ger**,**in**/out), BFW SmartInfo (**POI**,**Ger**,**in**/out)
Distance in steps	Sato et al. (2019a) (nav/**POI**,**in**), Nicolau et al. (2009b) (nav,**in**)
Direction as clock-face	Sato et al. (2019a) (nav/**POI**,**in**), Kamikubo et al. (2020) (nav,**in**), Blindsquare (**POI**,**Ger**,**in**/out), BFW Smartinfo (**POI**,**Ger**,**in**/out)
Direction as degrees	Kamikubo et al. (2020) (nav,**in**), Blindsquare (**POI**,**Ger**,**in**/out)
Direction as left-right	Paladugu et al. (2012) (**POI**,**in**), Kamikubo et al. (2020) (nav,**in**), Blindsquare (**POI**,**Ger**,**in**/out), BFW Smartinfo (**POI**,**Ger**,**in**/out)
Direction as cardinal points	Goyal et al. (2019) (nav/**POI**,**in**), Blindsquare (**POI**,**Ger**,**in**/out)

Table 4.3.: Overview of some of the rules for textual descriptions found in literature and existing Apps or prototypes. Note that some sources appear several times, as they feature a different grammar for POI than for navigation or for indoors than for outdoors, or they allow for different word order within the same context. In parentheses, we include information on whether the source features a navigation (nav) or POI context, if the instructions in the given source are in German (Ger), and if the context is outdoors (out) or indoors (in). The information that is most relevant to us appears in bold (**POI**, **Ger**, **in**).

was extended to be supported indoors as well. Indoors, however, there are two main problems: (1) map data for buildings is hardly available, and even then, the amount of information needed by people with visual impairments is not provided; and (2) a dense infrastructure of Bluetooth Beacons must be installed and maintained inside the buildings. Since we could not test BlindSquare ourselves in a building, we transcribed two videos, one in German and one in English, that show how BlindSquare works indoors (Neumann 2022). BFW Smartinfo is an Apple and Android app specifically designed for people with visual or mobility impairments, that offers information and orientation. It works both outdoors and indoors. It uses GPS, Bluetooth Beacons, QR codes and NFC tags. The QR codes and NFC tags are installed at fixed positions or on objects and can give additional, comprehensive information, for instance about a room or a coffee machine. RightHear is an Apple and Android app that works indoors, and is intended to be used in venues such as airports, shopping centers or office buildings. It gives only information about objects around. According to RightHear's official website, the app's main purpose is to provide 'talking signage' based on Bluetooth beacons. The audio descriptions associated with each beacon must be added manually in an online platform (RightHear 2018). Since we could not find out whether there are rules on how to provide the descriptions, or if they are completely manually inserted and thus different for each venue, we excluded RightHear from our analysis.

Table 4.3 shows the results of our analysis regarding: word order, amount of information, unit measure used for distance and how to output the direction.

Grammar Object-Direction-Distance:	([Details]* Object [Details]* Direction Distance)^n
Grammar Object-Distance-Direction:	([Details]* Object [Details]* Distance Direction)^n
Grammar Random:	Random permutations of “Object [Details]”, “Direction” and “Distance”

Table 4.4.: Main parts of the three grammar versions with different word order expressed as regular expressions: (1) Object-Direction-Distance (ObDirDis), (2) Object-Distance-Direction (ObDisDir) and (3) arbitrary word order (Random). Originally published in Constantinescu et al. (2022b).

4.2.2. Grammar Specification

The research papers and existing apps analyzed provide either navigation, or POI-based orientation, or both, in English or German, indoors, outdoors, or both. Thus, a conclusion of what was most appropriate in our context (POI-based orientation for indoors in German) was not obvious for all the details of the grammar, such as the word order or conveying directions. We thus decided to create several versions of the grammar and answer the remaining questions in a user study. In order to keep the user study manageable, we reduced the grammar versions to three:

1. **ObDirDis**: object followed by direction and distance,
2. **ObDisDir**: object followed by distance then direction,
3. **Random**: random word order.

The random version reflects the unstructured way in which people talk and give instructions. Asking passers-by for help is one of the strategies most often used by people with visual impairments when travelling. Thus, we use the random order variant of our grammar as a baseline. Table 4.4 shows the rules of the three different grammars, expressed as regular expressions.

Objects can contain details such as name or number of the room, *ladies'* or *men* in case of restrooms, reachable floors for stairs and elevators. In the example below, the object *staircase* has both a name, *TR06*, and additional information about where it leads to, namely *to floors 0-3*:

German: Treppen TR06 zu Stockwerken 0-3 auf 2 Uhr in 10 Metern.
English: Staircase TR06 to floors 0-3 at 2 o'clock in 10 meters.

The resulting descriptions must be as short as possible to allow users to focus on their environment and travel safely.

Table 4.5 shows an egocentric description of a building part generated with the three grammar versions. The order of the sentences depends on the position of the named object with respect to the user. Objects are thus listed clockwise, starting with 10 o'clock or 315° (front left or -45°). Column 2 is identical with column 1, except that Distance and Direction are interchanged. In column 3, the order of the three main components - object, direction and distance - is randomly chosen in each sentence.

German:

Grammar Object-Direction-Distance	**Grammar Object-Distance-Direction**	**Grammar Random**
Raum 109 auf 11 Uhr in 11 Metern.	Raum 109 in 11 Metern auf 11 Uhr.	Raum 109 auf 11 Uhr in 11 Metern.
Gang F02 auf 2 Uhr in 18 Metern.	Gang F02 in 18 Metern auf 2 Uhr.	In 18 Metern auf 2 Uhr Gang F02.
Haupteingang auf 6 Uhr in 1 Metern.	Haupteingang in 1 Metern auf 6 Uhr.	Haupteingang in 1 Metern auf 6 Uhr.

English translation:

Grammar Object-Direction-Distance	**Grammar Object-Distance-Direction**	**Grammar Random**
Room 109 at 11 o'clock in 11 meters.	Room 109 in 11 meters at 11 o'clock.	Room 109 at 11 o'clock in 11 meters.
Corridor F02 at 2 o'clock in 18 meters.	Corridor F02 in 18 meters at 2 o'clock.	In 18 meters at 2 o'clock corridor F02.
Main entrance at 6 o'clock in 1 meter.	Main entrance in 1 meter at 6 o'clock.	Main entrance in 1 meter at 6 o'clock.

Table 4.5.: Examples of egocentric descriptions inside a building generated with the three grammar versions: Object-Direction-Distance (ObDirDis), Object-Distance-Direction (ObDisDir) and random word order (Random), first the original in German followed by the English translation (Constantinescu et al. 2022b).

4.2.3. Evaluation of the Indoor Grammar

The main question to be answered in the evaluation of the grammar is whether egocentric textual descriptions created with our grammar support the orientation of users in unknown buildings (effectiveness). Moreover, we also assess the efficiency of the descriptions. Finally, we answer those questions that remained unclear in the grammar specification after analyzing the existing literature: order of words in the sentence, how to convey directions and the default radius.

Initially, we intended to evaluate the grammar in an on-site user study using the prototypical app MapEar. However, because of the Covid-19 restrictions at that time, but most importantly, due to our main intent of assessing the grammar and not the app itself (which would have added further variables to the user study), we decided to conduct an online evaluation. Later on, we also evaluate the grammar on-site, in a Wizzard-of-Oz study that simulates the output of the prototypical app, by comparing the textual descriptions generated with our grammar with sonification - see Section 5.3.

Before starting the questionnaire we analyzed previous literature for best practices in terms of evaluation methods. We wanted to see how others evaluated the interfaces of electronic travel aids in similar scenarios. Since research projects that deal with textual descriptions (2 projects) or maps (3 projects) are scarce, we included 8 projects that deal with navigation. Thus, 13 research papers were included in total in the analysis: Nicolau et al. (2009b), Kamikubo et al. (2020), Tröger et al. (2020), Sato et al. (2019a), Jain (2014a,b), Ko et al. (2011), Al-Khalifa and Al-Razgan (2016), Yang et al. (2011), Hersh (2020), Goyal et al. (2019), Madugalla et al. (2020), Kalaganis et al. (2021). In the case of navigation systems, the most popular evaluation metrics for the research projects analyzed are the deviation from the optimal route and the time needed to reach the destination. Since we do not feature navigation, both metrics are irrelevant to us. For descriptions in general, one method used is to compare to a baseline, while the metrics used are usefulness, the subjective experiences of the users and the cognitive load (Neumann 2022).

Based on this analysis and also on the research questions that we want to answer, we evaluated the effectiveness of the textual descriptions generated with our grammar in terms of the mental map that users develop. The baseline we use for comparison is the grammar version with random word order which reflects the way sighted passers-by give information when they are approached by people with visual impairments inside buildings. The efficiency is measured in terms of the cognitive load induced - a too high cognitive load makes the descriptions useless, as the user would not be able to travel safely anymore (to be considered that on-site, the cognitive load is expected to be even higher than in an online evaluation). We structure our questionnaire as follows:

Part 1. Demographics and questions about the mobility inside of buildings;

Part 2. Settings and information needs:

  - Questions on radius and how to convey directions;
  - Information needs: object categories;

Part 3. Usability: effectiveness, efficiency, perceived usability and usefulness:

  - Effectiveness and orientation indoors: mental map;
  - Efficiency: Cognitive Load (Raw TLX);
  - Usefulness and perceived usability (Net Promoter Score, short NPS);
  - Word order preference

Note that the last point evaluated, word order preference, actually belongs to the settings in Part 2, but can only be evaluated after the participants have read all three versions of the textual descriptions in Part 3.

We implement the online study in SoSciSurvey[5] running on local KIT servers. SoSciSurvey is a German, very powerful platform for creating professional online questionnaires with many customization options. But most importantly, the resulting questionnaires are accessible, unlike most other platforms. The questionnaire was designed and pretested in an iterative way with 3 experts in accessibility and assistive technologies, one of whom was blind. The goal was to assess the feasibility, comprehensibility and accessibility of the user study and to design the questions in such a way that we attain our goal without overstraining or frustrating the participants. As a result, many questions and texts were adjusted and the study was strongly shortened so that filling it in would last about one hour.

(1) In the first part of the questionnaire, after the demographics, we asked questions about mobility: the users' travel practices and needs, the use of digital navigation aids and the orientation strategies used in buildings. (2) In the second part, we asked about the preference for expressing directions indoors (as clock-face, left-right, degrees or cardinal points). We also provided a multiple choice list containing the objects implemented in our app, from which users had to choose only those objects that interest them. Users also had the possibility to suggest further objects and to mention, in a free text field, in what radius objects should be conveyed. (3) In the third part of the questionnaire, the three grammar versions (see 4.2.2) were used to generate German egocentric textual descriptions of three different buildings. Each description was limited to contain 10 objects each. The order of the grammars, as well

---

[5] https://www.soscisurvey.de/

as the combination grammar and building were randomized to avoid a learning effect and also to avoid that differences in the evaluation are due to the building's complexity instead of the grammar itself. In generating the descriptions, the directions were specified according to users' choice, mentioned in Part 2 of the questionnaire. After the user read the description generated with each of the three grammar versions, we asked questions about: mental map (describe the environment, questions about the location of specific objects), cognitive load (Raw TLX), perceived usability (NPS). The aim was to assess which word order is better and if it influences the cognitive load of the users. Also, we wanted to see if the users could build a correct mental map given the textual descriptions provided. (4) At the end of all three descriptions we asked further questions about word order preference, perceived usefulness, preference of textual descriptions over asking people on site or other preferred method. Finally, we analyzed the written descriptions of the environment made by participants in the section assessing the mental map.

The invitation to the study was sent via email to blind contacts of the ACCESS@KIT who previously consented to be contacted. We addressed the questionnaire at blind people only, as they are the intended users of textual descriptions. Visually impaired people with residual sight work most of the time visually, using the displayed map.

Participants. Eight blind people answered the first two parts of the questionnaire, and six of them completed the full questionnaire. The participants were between 18 and 67 years old. Four were male and four female, four were blind from birth.

Results of the indoor grammar evaluation. We present here the results of the evaluation, divided into several topics.

Orientation and mobility indoors. Traveling alone is an important aspect for people with blindness. Five of the participants would like to travel more on their own. Six users stated that this would bring them more independence. One user (P5) added that this in turn would bring "greater flexibility", while P4 added their "desire for self-determination". Only one person, P2, was satisfied with his current situation. He stated that he travels both alone and with attendance in buildings. No participant uses digital navigation aids inside of buildings.

Hearing and using natural guidance systems and cues such as walls, changes in the ground surface, sounds and smells is an important way to gather information about the surroundings. P4 and P6 for instance first stop and listen for distinctive noises that give them clues about the environment, such as from stairs, elevators, people. This points out at the importance of well designed auditory interfaces that do not interfere with the user's hearing of the environment. Our scenario of egocentric textual descriptions meets this requirement as the descriptions are called on demand by the user when needed.

Settings of direction and radius. Five participants prefer the clock-face system for conveying directions and three the left-right directions. According to P1, *"directions like right or back left are easier in theory, but you also have fewer subdivisions"*. No user has chosen degrees or cardinal points. Five participants do not consider the specification of the direction using cardinal points useful and two only conditionally.

When describing the environment in their own words, the 6 participants who finished the questionnaire used in the vast majority of their reproductions left-right directions, although 4 of them had previously chosen the clock system for conveying directions and only 2 left-right. Three participants, all of whom had chosen clock directions for the given descriptions, used left-right directions in their reproduction (36 times), clock directions (6 times) and both left-right and clock directions redundantly for the same object (6 times - P5,P6). The clock directions were thus used sparsely and often as a confirmation, in addition to the more natural left-right directions. The left-right directions, however, were not used consistently, not even within the descriptions of the same participant. The following example, listing expressions of direction from the participants' descriptions, illustrates why a standard wording is difficult:

> 1 o'clock: *very slightly to the right in front of me* (P1), *in front of me on the right* (P4), *diagonally to the right* (P4), *little further to the right* (P6);
> 2 o'clock: *a little further to the right* (P1), *on the right side* (P1), *next on the right* (P4), *on the more right side in front of me* (P4), *right* (P4), *further to the right* (P6), *on the right at 2:00 o'clock* (P6).

Three users, 2 of them only once each, expressed directions in relation to other objects: *further behind* (P3), *about there across the room [from the restrooms]* (P3). The prepositions *in* and *at* (German: "in", "auf") in front of the distance and direction entries were often left out by participants, which suggests that they could be left out in a grammar as well.

Regarding the radius within which to convey objects, the answers were very diverse: two users prefer $5m$ while the others stated: $2-20m$, $5-20m$, $20-30m$, "variable", and one wanted to have all prominent objects such as doors or stairs *in the room*, irrespective of the distance; alternatively, when looking for an object, either $3m$, or $5m$ for larger lobbies. Thus, for the default radius, we suggest using 20 meters and rather convey more objects than not enough.

Information needs. When asked what objects should be included in the descriptions, all participants selected *room*, *stairs* and *entrance* - see Table 4.6, while seven selected *doors* and *restrooms*. Each one selected between 7 and 15 objects out of 17, with an average of 11.5. Further objects named by the participants were: seating, cash/ticket counter, coat rack, room number, floor, wall, corridor and type of doors such as, for instance, glass door. Further, three participants wished to know more about the layout of the building and the current room. Thus, they wanted to have a very short description of the current space - wide corridor, open space (P1), to know the shape of the building (P3), get a rough structure - "overall description, info if there is corridor or similar in front of me" (P5).

Effectiveness: mental map. If the users can form a good mental map of the environment only from reading the descriptions, we can say that the egocentric textual descriptions generated with our grammar are effective - they achieve their goal. In order to assess the mental map, we asked users to read a description and then describe themselves, in written form, the environment. We analyzed the written descriptions by assigning 1 point for each correctly named object, 3 points for a correct object-direction combination, and -1 points for a false object. The grammar version Object-Direction-Distance obtained 137 points from all six participants, while the other two versions, Object-Distance-Direction and Random obtained each 132 points, namely on average 74.2% of the

Choice	Room	Stairs	Entrance	Toilets	Door	Corridor	Eating place	Open area	Info point	Shopping	Steps	Emergency exit	Tactile paving	Elevator	Ramp	Accessible toilets	Others
Count	8	8	8	7	7	6	6	6	6	6	5	5	5	4	2	2	1

Table 4.6.: Number of participants who selected the named objects.

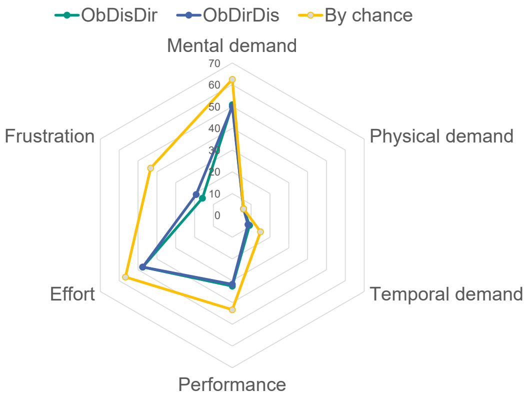


Figure 4.6.: Results of Raw NASA-TLX of the three grammar variants with different word order: (1) object-direction-distance (ObDirDis) (2) object-distance-direction (ObDisDir), (3) arbitrary word order (Random). Please note that the scale for performance is counter-intuitive, a higher score being worse (Constantinescu et al. 2022b).

maximum possible score. On average, each participant named 25.6 objects from 30 in total. Missing objects were usually those for which several instances of the same category were present in a single description. For instance, out of 3 stairs, only the closest one, or the first one mentioned in the description was referred to. The same could be observed for elevators and rooms. Also, most of the mistakes made by all 6 users usually involved these objects. Thus, another way of ordering objects in the descriptions could be to bundle all objects of one category together or to only display one object of each category - this, however, only if the objects are equivalent. Stairs or elevators that lead to different floors should all be listed.

The objects *wall*, *corridor* and *door* were mentioned by some participants, even though they did not appear in the initial descriptions. This suggests that information on rooms, taking the center of a room as reference point, is not enough: users need to know where the doors and walls are for orientation.

All descriptions were made from an egocentric point of view and most were made from the user's perspective: "in front of me there is...".. Only one participant used almost the same structure for describing the building part as in the given text. The other 5 participants used their own words and even included deduced elements which were not present in the given descriptions, including relations between objects (7 times - P1,P3,P5): "I stand in a *semicircular foyer*, [...] about there *on the opposite side* is the main entrance", "behind me is tactile paving *that presumably leads to* the stairs or elevators", "I am standing in a *long hallway*".

We conclude that the users could build a mental map with the egocentric descriptions provided, which prove to be *effective*, but there is also room for improvement, as the evaluation suggests.

Efficiency: cognitive load. We administered the raw NASA-TLX to assess the cognitive load of the users with respect to the three grammar variants. Using an arbitrary word order resulted in an increased cognitive load as can be seen in Figure 4.6. The users felt particularly frustrated when no fixed word order was used. Mental demand and effort are also slightly higher for the random sentences, while performance is lower (please note that the performance scale is inverted in the graph, a higher number meaning a lower performance). The other two grammar versions have almost identical scores. We thus conclude that egocentric textual descriptions generated by a structured grammar are more efficient (in terms of the cognitive load induced) than egocentric descriptions with random word order, as usually provided by natural persons - as one of the most popular strategies used by blind people when walking in unknown buildings is to ask other people for the way.

The mean cognitive load (RTLX) for the two fixed grammars, 27.7% and 27.5%, was low according to Grier (2015), for cognitive tasks. Thus, we could state that fixed descriptions are efficient and do not overload the user. However, when looking at the individual scores, mental demand and effort were quite high - between the 50th and 75th percentile for cognitive tasks (Grier 2015). For the grammar version with random word order, the mental demand was even higher, namely over 60, which is close to the maximum benchmark for cognitive tasks (Grier 2015). This implies that in a real navigation scenario, interpreting the textual descriptions with random word order takes a lot of mental effort, which might come at the expense of other cognitive tasks such as paying attention to the environment and traveling safely. As one participant in the study said, "descriptions have to be *structured* and *accurate* and reflect the information I need" (P4).

Usefulness and perceived usability. Five out of six participants found egocentric textual descriptions useful. One of the five conditionally: "but this depends strongly on the description" (P3).

We measured the perceived usability with the The Net Promoter Score (NPS). It is a short alternative to the System Usability Scale (SUS), consisting of one question only, with a 11-point Likert scale answer. The question we asked is: "Imagine you have a smartphone app that gives you information in the building, as given by the textual descriptions. How likely is it that you would recommend this system (from 0 unlikely to 10 very likely)?". In order to compute the score, the answers 0 through 6 are counted as detractors, 7 and 8 as passives and 9 and 10 as promoters. The score is computed by subtracting the percentage of detractors from the percentage of promoters. In our case, the two fixed grammar version had 2 promoters, 3 detractors and one passive each, while the random word order grammar version had 1 promoter only and 4 detractors. One participant only filled in the NPS for the fixed word order grammar variants, not for the random one. If we disregard this user, there is no great difference in the total NPS scores, as he was the passive user. The scores thus computed are -17% each for the fixed grammar orders and -40% for the random grammar. Among the reasons for dissatisfaction, the 3 detractors mentioned the large amount of information (P2) and the fact that one could not remember everything (P4). Another detractor wished to have a better overview of the environment's layout: "The method is not bad for getting an overview of what's there. What bothered me mostly is that there must be turns here, it can't all be at two o'clock, otherwise you'd have to walk through the toilets for example. The whole thing works well to derive a reasonable direction to a certain object, but I think I still would have liked more information about the layout of the place." (P1). One promoter (P5) also wished to be able to filter the object classes. The passive user (P6) said that textual descriptions greatly facilitate the navigation in large buildings, but thought that navigation instructions would be even better. For the random grammar variant, the inconsistent word order was criticized by two detractors and the amount of information by another. The promoter thought the random descriptions were very good, but wished the form of the room would be described (P3). The absolute values of the NPS assigned by participants were on average smallest for the random grammar (4.2 versus 6.8 and 7.2). For the second fixed grammar version, 2 participants gave better scores because they said they got better used to the descriptions, which points to the importance of learning a system. As a conclusion, users do not recommend the textual descriptions in the form provided in the evaluation, and wish instead they could filter the objects, set the radius and get an overview of the environment's layout, including the presence of walls - requirements most of which can be easily implemented in an app. Only the requirement regarding the presence of walls needs to be addressed in further research. Another conclusion is that a random word order results in a poorer recommendation (NPS score).

Word order preference. Besides testing the three grammar variants, we also asked users which word order they prefer from the three versions used in our grammar. Five out of six participants said they prefer the grammar variant Object-Direction-Distance, and one (P1) prefers Object-Distance-Direction. None chose the random word order variant. Participants P4, P6 and P1 said that the direction is more important than the distance: "first I want to know which object, so I know right away if I am searching for it or not, and then the direction, so I can place it in location" (P6); and "the most important thing is always the object, then where and how far, that's the best. Random output is very inconvenient" (P4). Participant P1 chose to have the direction at the end and argues that "this description type concludes with the most important information (the direction). If numbers are announced afterwards (the distance), then it can happen that one does not pay enough attention to the direction". Thus, although the three participants agreed that direction is more important than distance, the preference for its position in the sentence was different. This points at the importance of allowing the user to set the word order individually.

The analysis of the written descriptions of the environment made by participants based on the given textual descriptions also yielded interesting findings with respect to the word order. In the given descriptions, the word order was either: Object-Direction-Distance, Object-Distance-Direction or random. Participants, however, began 50 out of 94 times in total with the direction. In 40 out of these 50 cases, the direction was followed by the object only. In 36 cases, sentences from participants began with the object, and only 4 times with the distance. The distance was only present in 37 sentences out of 94. If at all present, distances were often imprecise or implicit, i.e. expressed in relation to other objects and to one's own position in the room: *further, closer, before, after.* Only 2 participants provided explicit distance information, one of them just a single time. Finally, out of 37 distances mentioned, 6 were false.

Order of objects. The order in which objects were mentioned in the descriptions made by the participants was most of the times similar to the given descriptions, namely clock-wise, beginning left, and in the same direction sorted by distance. In 5 out of 18 cases, three participants mentioned the entrance right in the beginning, although it never occurred first in the given descriptions. However, one participant did not mention the main entrance at all twice (from 18 instances in total, for all participants), which could indicate that for this person, the entrance is not very important, or at least not at the given time. In 3 cases, 2 participants seemed to order the features according to their importance, for instance: *entrance - stairs - toilets - rooms* (P5); *entrance - stairs - corridor - toilets - rooms*; *entrance - elevators - wall - room - staircase - rooms* (P3). One participant (P6) mentioned the objects in counter-clockwise direction. Objects were often grouped together, either by direction (12 times) or by object category (4 times): *On my right(3x) comes a staircase and a room E001* (P4); *On the left and right there are* 5 *rooms* (P3).

Discussion.

Entrance and choice of objects. The (main) entrance seems to be the most important object, which also suggests its possible role as a point of reference in allocentric descriptions. Other prominent features mentioned by participants, which should be present in the description, are corridors and walls, which give an overview of the space structure. All other features seem to be a question of personal preference.

Conveying directions. Clock directions are more precise than left-right directions, as they have 12 divisions as opposed to 8 for left-right. Left-right directions, on the other hand, are more natural, and it could be that some users cannot interpret clock directions. The participants used most of the time left-right in their descriptions in order to convey directions. However, even within descriptions of the same participant there is a great amount of variation and inconsistency in the usage of language. Thus, the question of what system to use for conveying directions (clock-face or left-right) could be one of the few configuration questions asked before the first use of the system.

Intuitive user interfaces. One participant said he would like to use digital navigation aids, but this “requires a lot of familiarization and [he does not] find the time for it, unfortunately” (P4). This points out at the importance of intuitive user interfaces and well thought default settings for an easy start and a straightforward way of changing settings afterwards according to own preferences. Customization before the first use of the app is not ideal, as users don’t know what they want until they have tried it (see also the evaluation of the outdoor grammar in section 4.1.3). Moreover, a lengthy set up discourages users, while using right away a product with good results encourages further use.

Recommendations. Based on the evaluation, we can make the following recommendations for an indoor grammar as well as in general for the systems implementing it:

- Use fixed word order in sentences.
- Provide the option to change the order of words in the sentence.
- Add an allocentric overview including a description of the form and structure of the environment before the list of objects.
- Provide as many objects as possible and allow users to filter them according to their needs and preferences. By default, at least entrance, stairs, toilets, rooms and doors should be activated. Walls and corridors should also be present in the description by default, whether as objects or in another form.
- Both clock directions and left-right should be provided as options.

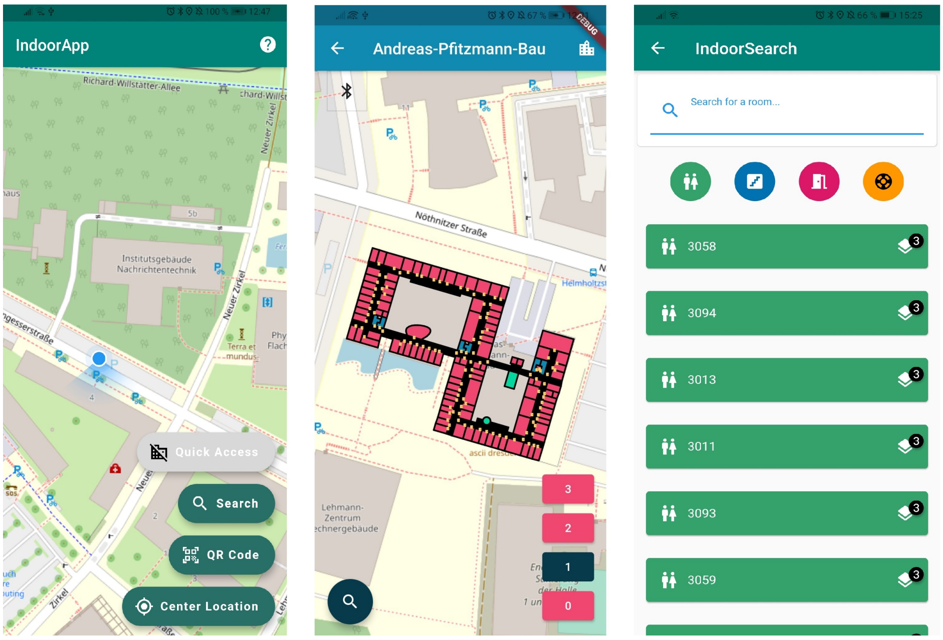


Figure 4.7.: The main functionality of MapEar as implemented by TU Dresden (Neumann 2022). Left: outdoor map view. In order to view a building, the user can either click on a mapped building or search for its name. Middle: Indoor map view with one button for searching inside the building and buttons for switching between levels. Right: Indoor search view.

- Provide an option to include/exclude distance information in and from the descriptions, as distance is less important than object and direction.
- The *default* radius in which to convey objects is 20 meters, with the possibility to change it from options.

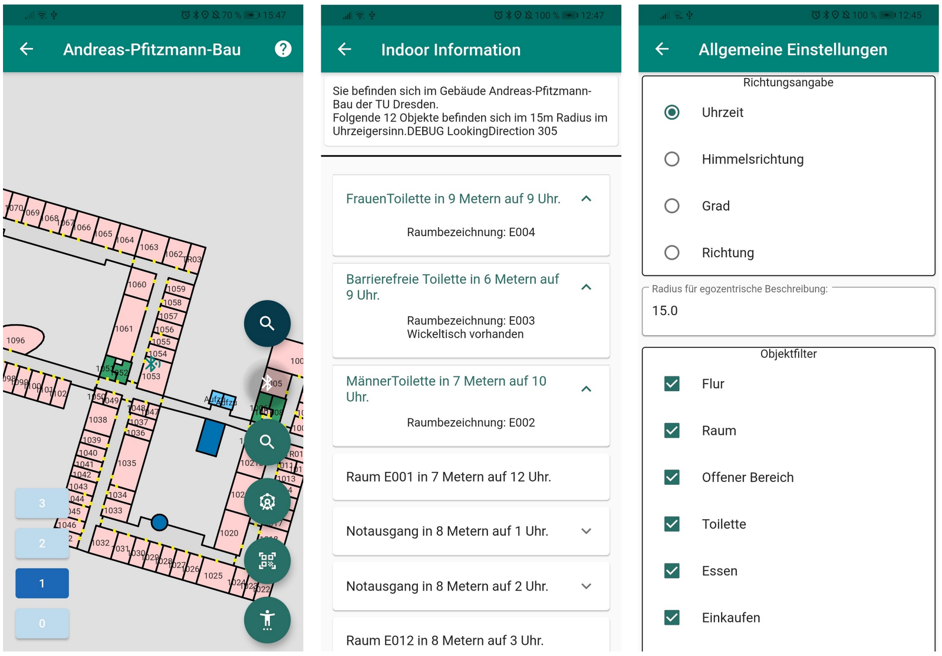
Complementary to the online user study, we also conducted an on-site evaluation inside a building, as a comparison between speech and sonification - see section 5.3.

4.2.4. Proof-of-Concept: Prototypical Implementation of the Indoor Grammar

As a proof-of-concept, we integrated the grammar into an existing prototypical app, provided to us by the AccessibleMaps[6] project partners from the Technische Universität Dresden (TUD). The app, which we call MapEar, is optimized for Android. The prototypical app developed by TUD retrieves selected building data from OpenStreetMap and displays it on a visual map. The visual map is intended for sighted escorts of blind people or for visually impaired people with enough residual sight. The app is otherwise accessible to blind people and is specifically designed taking screen reader usage into account. The programming is done in the Dart[7] programming language under the Flutter[8] framework. Both Dart and Flutter were created by Google to ease the development of multi-platform apps. The retrieval of OSM data is done through the Overpass API via Microservices. Besides the display of

---

[6] https://accessiblemaps.de/
[7] https://dart.dev/
[8] https://flutter.dev/



Sie befinden sich im Gebäude Andreas-Pfitzmann-Bau der TU Dresden.
Folgende 12 Objekte befinden sich im 15m Radius im Uhrzeigersinn.DEBUG LookingDirection 305

FrauenToilette in 9 Metern auf 9 Uhr.
Raumbezeichnung: E004

Barrierefreie Toilette in 6 Metern auf 9 Uhr.
Raumbezeichnung: E003
Wickeltisch vorhanden

MännerToilette in 7 Metern auf 10 Uhr.
Raumbezeichnung: E002

Raum E001 in 7 Metern auf 12 Uhr.

Notausgang in 8 Metern auf 1 Uhr.

Notausgang in 8 Metern auf 2 Uhr.

Raum E012 in 8 Metern auf 3 Uhr.

Richtungsangabe
● Uhrzeit
○ Himmelsrichtung
○ Grad
○ Richtung

Radius für egozentrische Beschreibung: 15.0

Objektfilter
☑ Flur
☑ Raum
☑ Offener Bereich
☑ Toilette
☑ Essen
☑ Einkaufen

Figure 4.8.: The implementation of our grammar in MapEar as egocentric textual descriptions (Neumann 2022). Left: indoor map view with the two additional buttons for egocentric textual descriptions and their settings. Middle: Example of an egocentric textual description generated with our grammar. Right: Settings of the egocentric textual descriptions.

map data, the MapEar app was developed by the TU Dresden to also detect the location of the user (namely of the mobile device held by the user), either from Bluetooth Beacons or QR-Codes. We decided to use Bluetooth Beacons, since QR-Codes installed in a building must be captured by users with the mobile device's camera, which is a very difficult task for blind people. In Figure 4.7 one can see the functionality initially provided by the app before we integrated our grammar.

We extended MapEar by an egocentric description component based on our grammar, or, in other words, textual descriptions of the indoor environment from the user's perspective, taking into account the position inside the building and the direction in which the mobile device is heading when held in the hand. Two additional buttons in the indoor map view allow the user to: (1) call the egocentric textual descriptions and (2) change the settings of the descriptions, namely: choose the categories of objects to be conveyed, specify how to convey the directions (as clock-face, left-right, degree or cardinal points) and set the radius in which to convey objects. Figure 4.8 shows the added functionality. As default settings, all object categories are selected, the radius is set to 20 meters (in the image it is changed to 15 meters) and the directions are output as clock-face. There are 17 object categories supported, namely: corridor, room, open area, restroom with sub-category barrier-free restroom, food facility, shopping facility, info point, elevator, stairs, steps, ramp, door with sub-categories entrance and emergency exit, tactile paving and miscellaneous. These object categories were chosen taking into account both users' needs, namely the indoor accessibility features most important to users (see Section 3.1 as well as the results of our evaluation in the previous Subsection, 4.2.3) and the technological affordances, namely the objects that can be mapped in OSM.

When the user presses the egocentric description button, the location and orientation of the mobile device is determined using Bluetooth Beacons and the mobile device's compass. This information is then saved and used for generating the description. Neither the location nor the orientation are updated anymore once the button has been pressed. This is very important for blind people, otherwise they might get confused if the description changes dynamically. Thus, it is assumed that the user stops and holds the mobile device still until all the needed information is read out.

The egocentric descriptions begin with a short meta-description of the environment containing the name of the building and the number of objects found given the settings. After that the list of objects is displayed (and read out, respectively) clock-wise, starting front left. For some objects, additional information is available, such as, for instance, the room number, the door type or whether a diaper changing table is available in a restroom. This information can be blended in under the object's main description by expanding it - action which is indicated by a small arrow on the right (see Figure 4.8, the image in the middle).

The implementation of the egocentric descriptions in the prototypical app MapEar was done in a user-centered approach, with the involvement of three accessibility experts, one of whom was blind. The experts evaluated the first versions of the prototype and its functionality and gave feedback that was taken into account in subsequent iterations. A thorough evaluation of the prototype with users is still to be done in future work, but this implies solving the problem of indoor localization for people with visual impairments first. Our recommendation is to offer the app for free to users and conduct a long-term evaluation with log data gathering and user questionnaires.

4.2.5. Conclusion

In a user-centered approach, we designed a German grammar for conveying information about objects in the environment inside a building. In an online evaluation with 8 blind people, we showed that our grammar is *effective* in helping users form a mental map of the described environment. Users need and wish a structured grammar with fixed word order. A random word order frustrates users, resulting in higher cognitive load and in lower *efficiency*. Based on the evaluation, we provide recommendations for indoor POI-based grammars and the systems implementing them. Finally, we implemented our grammar, in form of egocentric textual descriptions of the environment, into a prototypical Flutter app as a proof-of-concept.

5. Non-Speech Auditory Displays

While speech-based auditory displays are currently the state-of-the-art in output interfaces for people with visual impairments, speech also has its limitations. For some users, for instance, speech is too slow, as we have seen in section 4.1.3. Another issue is that speech is language-dependent, excluding thus people who do not speak well enough a certain language. The current migration situation worldwide encourages us to research an alternative interface, namely non-speech sonification. The question addressed in this chapter is *how* to convey information to people with visual impairments in the context of travel using non-speech sonification. In this chapter, we analyze auditory icons combined with parameter mapping for conveying information about objects in the environment. Among the sonification techniques (Hermann et al. 2011), auditory icons are suitable for conveying information about the different objects (their names) in an intuitive way. For conveying the distance and position with respect to the user, we use parameter mapping. That means changing sound parameters such as volume, pitch, panning, etc. to convey information about distance and position. The challenge is to create auditory icons that are short enough to be faster than speech. One advantage is already won through the parallel communication of object class, distance and direction. But, given that people with visual impairments can use speech at high speeds[1], keeping the auditory icons as short as possible is paramount for providing an advantage in terms of speed over speech. In section 5.1 we implement an interface based on short auditory icons with parameter mapping into a prototype and evaluate it with target users in an on-site user study as a proof-of-concept. The goal is to assess if this approach is appropriate for people with visual impairments in travel scenarios. Since the outcome is positive, in section 5.2 we create, in a user centered approach, a set of very short auditory icons for about 40 everyday objects and evaluate them in an online questionnaire. Finally, in section 5.3 we compare speech and sonification in an on-site user study inside a building.

5.1. Preliminary Study: Outdoor Object Sonification

This section is based on our ICMI publication Constantinescu et al. (2020).

As already mentioned, besides navigation information, people with visual impairments also need information about objects around them, whether POIs, landmarks or obstacles. Mobile solutions based on map data can provide some information about few static “objects” such as certain types of buildings, traffic lights or crossings. However, this is by far not enough. People with visual impairments also need to know about bicycles parked on the sidewalk, cars, approaching people,

---

[1] Research suggests that “the best compromise between efficiency and the ability to understand each sentence is the use of [...] a rate of 1.75*default-rate (approximately 278 WPM)” (Guerreiro and Gonçalves 2015).

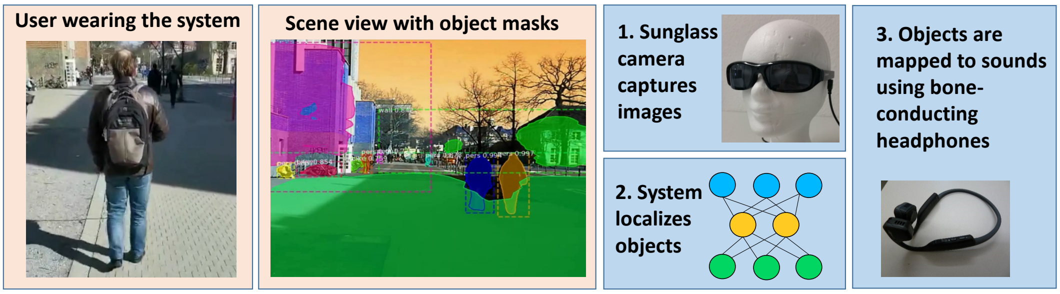


Figure 5.1.: Overview of our three-step process used in the outdoor object sonification prototype: (1) a camera captures images of the environment; (2) the system localizes objects of the environment using a deep learning neural network; (3) the user is informed about the recognized objects via sonification (Constantinescu et al. 2020).

benches to sit on and many other dynamic objects that cannot be mapped into geographic databases. Recent advances in computer vision and artificial intelligence makes it possible to recognize all these objects. We propose a prototype for supporting people with visual impairments during travel by recognizing both static and dynamic objects and conveying the recognized semantic information to the user via sonification (see Figure 5.1). We evaluate the prototype in an on-site user study with five people with visual impairments. In this context, we address several questions, with the primary goal of assessing the suitability of the sonification method for the given context and target user group: (1) *Is the output intuitive*[2] *and, thus, easy to learn by the target group?* (2) *Are users satisfied with the system in general and with the interface in particular (user satisfaction)?* and (3) *What are the users' preferred settings for object classes, radius and angle?*

Sonification is being used in various applications and research projects for people with visual impairments, from access to mathematics in general and graphs in particular (Walker and Mauney 2010, Vines et al. 2019) to support in various navigation tasks (Ahmetovic et al. 2019, Wilson et al. 2007, Katz et al. 2012, Ziemer and Schultheis 2018, Aziz et al. 2019) or for orientation (Geluid in Zicht Foundation 2014). Few of these approaches, however, leverage auditory icons. This sonification method was first used in computer systems in the 1980s (Gaver 1986) and later in both mainstream and audio games (Impression Games 2020, GMA Games 2020). Literature on this subject is scarce, especially with the focus on people with visual impairments. We list here a few of the research papers most relevant to us.

Ferati et al. (2011) propose audemes, which are based on auditory icons, to facilitate interaction with large collections of spoken educational essays for pupils with visual impairments. However, in contrast to our system, they do not apply sonification to convey objects in real-time scenarios, and the design of the auditory icons does not consider the duration of the sounds.

Aziz et al. (2019) explore several sonification methods for auditory routes overviews, including text-to-speech, earcons, and auditory icons. They conclude that auditory icons are appropriate to convey information about points of interest. While they do mention that the auditory icons should be intuitive, they also suggest that a training phase should be provided to help users understand the mapping scheme.

---

[2] Please note that "intuitive" is used here to mean "directly apprehended" or "readily [...] understood" (Dictionary by Merriam-Webster. America's most-trusted online dictionary 2020)

Tislar et al. (2018) investigate sonifications of objects through music, earcons, spearcons and lyricons regarding their learnability, the relatedness of sounds, their attributed meanings as well as their intuitiveness. However, they do not include auditory icons and do not evaluate with visually impaired people. Thus, their results are not transferable, as people with visual impairments and sighted people have different preferences for user interfaces (Walker and Mauney 2010), as well as different cognitive loads when using them (Martinez et al. 2014).

Dingler et al. (2008) propose a work most similar to ours in terms of sonification. They compare, in terms of learnability, auditory icons, earcons, spearcons and speech used for representing objects. But this work does not mention any limitation of the *duration of the sounds*, which is relevant to present various objects in a short time. They also trained the sounds first, while we investigate the intuitiveness first. Finally, they exclusively evaluated their system with sighted students. In Biggs et al. (2019), short auditory icons (up to 700 ms) were used, but no evaluation of the sounds or comparison is described in the paper.

Some approaches also use computer vision to detect objects in indoor or outdoor environments informing the user via speech (Hub et al. 2003, OrCam Technologies Ltd. 2020, Bashiri et al. 2018) or vibrotactile feedback (Zeng et al. 2017). Recent systems combine speech with other computer vision techniques such as: generating a textual description of the entire scene (Microsoft 2020) or predicting a list of object classes (Aipoly Inc 2020). A disadvantage of these approaches is the use of a phone camera, which is very difficult for people with blindness to hold straight and also impractical when already holding a white cane in the dominant hand (Saha et al. 2019). In comparison, our system continuously captures the entire scene, conveys both near and far objects to the user, is hands-free and does not require the user to hold the camera straight trying to capture an object or area of interest.

Other related methods try to incorporate a combination of computer vision and non-speech sounds to help the orientation of people in indoor environments (Ribeiro et al. 2012), to identify known people in the environment through face recognition techniques, to help users find close objects using sonification methods (Nagarajan et al. 2003, Auvray et al. 2007, Sainarayanan et al. 2007, Tang and Li 2014, Schauerte et al. 2012), to warn cyclists about objects not in their field of vision (Schoop et al. 2018), to give guidance during road crossing (Mascetti et al. 2016) or to warn with beeping sounds when other people block the way (Kayukawa et al. 2019).

The most relevant approaches are the works proposed by Katz et al. (2012) and Presti et al. (2019). Katz et al. (2012) use computer vision methods to detect objects in outdoor environments, which are then sonified using 3D sound. The main difference to our approach is that their system can only localize a *single* class of objects at once upon user's request. So basically, the user is searching for a certain object. In comparison, we can localize up to 18 different object classes *simultaneously* and sonify them using several different sounds. Presti et al. (2019) developed an iOS application to detect obstacles. If several obstacles are detected, the system informs the user about only one of them by transmitting its distance, size and direction via a combination of a base sound and an auditory icon. In comparison to this approach, our system does not only localize different objects in the scene but also *recognizes* them. People with visual impairments can use our system both for (1) orientation or macro navigation (Strothotte et al. 1996) by hearing static objects like walls, traffic lights, benches, and for

Figure 5.2.: Hardware of the sonification prototype without iPhone: backpack with laptop, sunglasses with small integrated camera, bone conducting headphones.

(2) obstacle avoidance or micro navigation, by hearing static and dynamic objects at close range like persons or bicycles.

The hardware of the prototype (see Figure 5.2) is similar to the prototype in subsection 4.1.3 and consists of a smartphone (iPhone), a laptop in a backpack, a camera, this time integrated into sunglasses, and Bluetooth bone conduction headphones - as first suggested by Wilson et al. (2007). The software of our system includes a computer vision module and a sonification module. The smartphone app (Ritterbusch et al. 2018) is the core of the prototype and connects the laptop with the headphones, handles the sonification and controls the user settings. Figure 5.1 summarizes the interaction between the different components: the camera from the sunglasses captures objects from the environment and transmits the frames through a short cable that runs behind the user's ear to the computer vision module. The objects are identified and the information about object class, distance and direction is sent to the sonification module. Here, the information is mapped to sounds and passed on per Bluetooth to the bone conduction headphones (AfterShokz Trekz Titanium) worn by the user. In one case, the headphones were not used, as the hearing aid of a participant was directly connected to the smartphone.

The list of objects that the prototype identifies was obtained through the intersection of users' needs and technological affordances. The users' needs were collected in a previous requirements analysis conducted by Dr. Karin Müller. By technology affordances we mean the object classes available in the datasets used, namely COCO (Lin et al. 2014) and COCO-Stuff (Caesar et al. 2018). Our object list thus obtained contains 18 object classes representing static and dynamic objects that can be found in urban environments (see Table 5.1). The neural network based computer vision algorithms were implemented by Dr. Monica Zündorf, born Haurilet (Constantinescu et al. 2020).

For sonifying the information about objects obtained from the computer vision algorithm - object class, direction, distance - we use parametric auditory icons (Schito 2012), meaning a combination of auditory icons with parameter mapping. In our case this means that each object class has its own auditory icon. Additionally, we map object data to sound parameters: the object's distance to the user (camera) maps to loudness and the x-axis displacement of the object with respect to the user maps to panning. Panning means that the sound seems to come from the direction of the object. We thus use one auditory icon for one object class and set the parameters in real time for the specific object instance. Mapping distance to volume has the advantage that far away objects, which are less important

Object	**Conceptual mapping**	**Sounds duration (ms)** Auditory Icon	Speech	**German word**
Bench	Person sitting down on a creaky wooden bench	456	806	Parkbank
Bicycle	Sound made by bicycle when rolling without pedalling	366	657	Fahrrad
Bus	Bus engine	432	525	Bus
Bush	Rustling leaves in a bush	366	525	Busch
Car	Car engine	648	546	Auto
Chair	Person sitting down on a creaky chair	672	612	Stuhl
Dog	Dog barking	240	587	Hund
Door	Knocking twice at a wooden door	313	556	Tür
Fence	A wire fence being kicked	705	692	Zaun
Motorcycle	Motorcycle engine when firing up and accelerating	575	812	Motorrad
Person	Person walking 2 steps on gravel	480	763	Person
Stairs	Walking 2 steps down on a wooden stair	552	717	Treppen
Stop Sign	Beep for traffic sign from navigation app or car info system	552	860	Stoppschild
Traffic Light	Modified sound of accessible traffic light with acoustical signal	1228	640	Ampel
Train	Train whistle	392	610	Zug
Train Tracks	Signal bell at railway crossing, heavy train on background	360	628	Gleise
Truck	Truck motor on background, beeping sound when driving reverse	840	767	LKW
Wall	White cane hitting twice against a wall	504	582	Wand

Table 5.1.: Conceptual mappings of the auditory icons for 18 object classes, the sound duration of the auditory icons and of the respective speech sounds and the German words used for generating the speech. The German words were generated with a text-to-speech engine with a speed of 175 words per minute.

or less urgent, are more quiet and thus do not mask environmental sounds. We only take into account the position of the object on the x-axis. We do not consider the y-axis, or height, as the difference on the y-axis cannot be perceived on bone conducting headphones. We chose auditory icons combined with these parameters (panning for x-displacement and volume for distance) due to their intuitiveness - see requirement R11 in Table 3.5. We considered that the users should be able to use the system, as much as possible, right away, and not have to learn the interface first. Only this way the interface can be comparable to and compete with a speech interface, which is currently the standard and most intuitive. To create the auditory icons, we chose sounds that are intuitive and can be easily matched with an object. For instance, a car is sonified by a car's engine and a bush by the rustling leaves in a bush. Table 5.1 lists the conceptual mappings of all 18 object classes.

The sounds must be very short in order to provide an advantage over speech, but also because there can be many objects in one frame that must be played sequentially. This is a difficult task as sounds loose their recognizability very quickly when shortened. In the end, the duration of 17 out of the 18 sounds ranged between 240 ms for dog and 850 ms for truck. Alone the sound for traffic light exceeded 1 second and lasted 1240 ms. We compared the duration of the auditory icons in milliseconds (ms) with the duration of the corresponding spoken words in German at default speech rate.

We generated the speech with Gespeaker (Castelli 2020), a free GTK+ frontend for espeak (Free Software Foundation, Inc. 2020), using the German-mbrola-5 voice at the default speed of 175 words per minute and a delay of zero. The average play time for all 18 sounds was 538 ms for the auditory icons and 660 ms for speech - which is 23% longer. The difference is 122 ms and the standard deviation (SD) 230 ms. The difference is statistically significant, with $p = 0.037$ in a paired t-test. When omitting the sound for traffic light because of its exceptional long duration, the average difference between auditory icons and speech is 164 ms (speech on average 33% longer), $SD = 151$ ms and $p = 0.00038$ in a paired t-test. This shows that the chosen auditory icons are shorter than speech for the given objects. Please note that we chose the shortest, most representative words for naming the objects, deliberately avoiding long terms. Additionally, if we also need to convey other information about the object besides its class, such as position, distance or size, for instance, the speech becomes longer, while the parameterized auditory icons remain the same length. As an example, a speech instruction containing the object class, the direction expressed as clock-face and the distance, could be: "car at one o'clock in two meters". Or, a more shortened version: "car at one in two meters" (German: "Auto um eins in zwei Metern"). The duration of the German instruction at the standard speech rate of 175 wpm is 1555 ms, while the auditory icon for car only takes 648 ms to play. Coincidentally, at 500 wpm, which is the maximum speech rate used by advanced users according to Edwards (1991), the speech output for this phrase is equal to the length of the auditory icon, namely 648 ms. It will have to be tried, though, for any given set of objects, which of the two modalities is on average faster and at what speech rate auditory icons offer an advantage over text-to-speech in terms of speed.

Another advantage of keeping the auditory icons as short as possible is the fact that they do not resemble as much the original, natural sounds. Thus, they are easier to distinguish from environmental sounds, which is a very important aspect when using the auditory icons outdoors.

In our prototype, the objects found in the environment are conveyed sorted by distance, staring with the closest. This is because the closest objects are usually the most important: they might be obstacles that the user needs to avoid, or a free bench that the user is looking for. The user can turn objects on and off at any time from the options, and can also set the radius (maximum distance) within which objects should be sonified. We set the default distances to be different for different object categories, according to their speed: for fast vehicles (cars, motorcycles, buses, trucks, and trains) it is 30 meters, for bicycles 15 meters, for people and dogs 10 meters and for static objects also 10 meters. For example: if only bicycles are checked in the options and the radius is set to the default of 15 meters, then only information about bicycles within 15 meters from the camera is transmitted to the user. The larger radius for vehicles and bicycles is necessary to give the system enough time to identify and convey information about fast moving objects to the user, and to give the user time to react, if necessary.

Before the evaluation, we carried out a pretest with two accessibility experts and adjusted the prototype according to the insights obtained. We then evaluated our sonification interface and the entire prototype in an exploratory study with five people with visual impairments. The goal of the evaluation is to investigate whether and how a system that localizes objects and conveys them via sonification can be useful to people with visual impairments. We do this by addressing the following questions:

(1) How to design the interface to meet the users' needs?

(2) Are auditory icons suitable to convey information about objects in a scene?

(3) Are the conceptual mappings of objects to sounds appropriate and do they facilitate learning the auditory icons?

(4) Which objects, including obstacles, points of interest and orientation landmarks, are most important for people with visual impairments?

Participants' demographics. Five male adults with ages ranging from 21 to 50 years and an average age of 30.2 years (SD 11.7) participated in the study. One of them (P1) was severely visually impaired and could see objects at $10 - 20$ m ahead, depending on the light conditions. Two (P2, P4) were congenitally blind and could not see at all, and two (P3, P5) were legally blind according to German law, one of whom (P3) could see objects at very close range, according to his own statements. All of them travelled most of the time alone and on foot and four also used public transportation besides walking. Three out of five (P3-P5) used digital navigation aids, one (P2) very rarely and one (P1) not at all. All participants had received mobility training before. Four of them had a white cane and used it during the evaluation. One (P3) had neither a white cane nor a guide dog. An overview of the participants is given in Table 5.2.

We started our study by explaining to the participants the main idea of the system and the procedure. The subjects then signed the statement of consent and filled in a questionnaire on demographics. The study was divided into four parts: sounds evaluation, outdoor evaluation, assessment (filling in questionnaires) and comparison of two sonification methods. Only the second part, the outdoor evaluation, was conducted on-site, while all others were conducted in the laboratory. The duration of the entire study was around $2 - 3$ hours per participant.

(1) Sound evaluation phase. The participants were first accustomed to the auditory icons, while we also tested the intuitiveness and learnability of the sounds at the same time. In a first step, the test leader read aloud the list of 18 objects. Then, the sounds for all objects were played in a random order, and after each sound, the participants were asked to guess the object it represented (testing the intuitiveness). If the participant was not able to guess it correctly, the test leader told them the object name as well as the conceptual mapping (how the sound was created and how it relates to the object). Any sound could be repeated upon user's request. In a second step, the sounds were played again in a random order to test if the participants could remember them correctly.

(2) Outdoor evaluation phase. After the evaluation of the sounds, we asked the participants to walk on an urban route familiar to them and think-aloud (Nielsen 1994) while walking. With this evaluation, we wanted to learn how people with visual impairments use acoustic information about the outdoor environment. The test leader accompanied the participants, ensuring their safety and taking notes. The session was recorded with an action camera carried by the participants and attached to the backpack strap.

(3) Assessment phase. After the walk, the participants were asked to fill in a questionnaire that was partially based on the "I like, I wish, what if" method (Interaction Design Foundation 2020). We asked them: (1) what they liked, (2) what they did not like, (3) suggestions that may not have a link to the prototype; and additionally, (4) how useful they found each of the 18 object classes (on a Likert scale rating from 1 to 5) and in what situation, (5) to name three other relevant objects not included in

ID	Age Span (Years)	Gender	Visual Impairment	Onset	See Objects?
P1	$30-39$	m	low vision	childhood	Yes, depending on light
P2	$20-29$	m	blind	birth	No
P3	$20-29$	m	blind by German law[3]	age $20-29$	Yes, in front
P4	$50-59$	m	blind	birth	No
P5	$20-29$	m	blind by German law	after childhood	No

Table 5.2.: Demographic data of the participants in the preliminary study of the outdoor objects sonification (Constantinescu et al. 2020).

the test, (6) at what distance and (7) at which angle should objects be identified, (8) if they liked the camera integrated in the glasses, and (9) if and in what situations they would use the system.

(4) Comparison of sonification methods phase. Although auditory icons were our main interest, we also wanted to see how they compare to spearcons in terms of intuitiveness and users' subjective assessment (like/dislike). Spearcons are text-to-speech snippets played at very high speeds, to the point where the initial words cannot be understood anymore. Since some blind people use their screen readers with very high speech rates, this interface could be a good alternative to auditory icons. At the end of the study, we tested the intuitiveness of spearcons in the same way as for auditory icons. We played the spearcons in random order and asked the users to guess which object it represents. Users were also asked which approach they prefer, auditory icons or spearcons.

We evaluated our study by analyzing (1) the times and success rates from the sound evaluation phase and (2) the questionnaire together with the comments made by the participants during their walk with the prototype. From the questionnaire, we wanted to find out what objects in general are most relevant for people with visual impairments during navigation. Additionally, we wanted to know how the users intend to use the interface regarding the number of objects and the maximum distance at which objects are announced.

Intuitiveness and learnability of the object-to-sound mapping. We evaluated the intuitiveness of the auditory icons by analyzing the success rates of the participants in guessing which sound belongs to which of the objects. In a second round, we evaluated the learnability rate by assessing how easily the participants remembered the object classes associated with the sounds, as proposed by Hermann et al. (2011).

The values of the rating were as follows:

1 - if the user did not guess or did not remember the meaning of the sound

2 - if the user guessed or remembered the sound after thinking for longer than 3 seconds

3 - if the user guessed or remembered the sound within less than 3 seconds

Table 5.3a shows the average intuitiveness rates for all five participants (note that we are missing 3 values for intuitiveness and 2 values for learnability, from 180 values in total). **The most intuitive sounds were "bicycle", "door", "bush", "dog"** and **"motorcycle"**. The least intuitive sounds were

	Intuitiveness		**Learnability**	
Objects	Average	SD	Average	SD
Bicycle	2.8	0.45	2.4	0.89
Door	2.6	0.55	2.8	0.45
Bush	2.4	0.89	2.4	0.8
Dog	2.4	0.89	2.8	0.45
Motorcycle	2.4	0.89	2.6	0.55
Bus	2	0.71	2.2	0.84
Train	2	0	2.8	0.45
Fence	2	1	2.4	0.89
Traffic Light	1.8	0.84	2.6	0.55
Train Tracks	1.6	0.89	2.6	0.55
Car	1.4	0.89	2.6	0.55
Person	1.4	0.98	2.8	0.50
Stairs	1.4	0.89	2.6	0.55
Chair	1.4	0.89	2.8	0.45
Wall	1.2	0.45	2.6	0.55
Truck	1	0	3.0	0
Stop Sign	1	0	2.8	0.50
Bench	1	0	2.2	0.84
Average	1.8	0.6	2.6	0.6

(a) Intuitiveness and learnability, on a scale from 1=poor to 3=good, sorted descending by intuitiveness average (Constantinescu et al. 2020).

	Importance	
Objects	Average	SD
Stairs	4.8	0.45
Traffic Light	4.8	0.45
Bicycle	4.6	0.89
Motorcycle	4.4	0.89
Bench	4.2	0.84
Truck	4.2	1.10
Door	4.0	1.22
Fence	4.0	0.71
Car	3.8	1.30
Train Tracks	3.8	1.10
Train	3.6	0.89
Chair	3.4	1.67
Bus	3.2	1.30
Person	3.2	1.10
Wall	3.2	1.30
Dog	2.8	0.84
Bush	2.4	1.14
Stop Sign	1.6	1.34
Average	3.7	1.0

(b) Ranking of the importance of objects from 1=useless to 5=useful, in descending order.

Table 5.3.: Intuitiveness, learnability and importance ranking of the auditory icons evaluated, with average for all five users and standard deviation (SD).

for “truck”, “stop sign” and “bench”. “Truck”, however, had the highest learnability rate. This means that once the conceptual mapping was known, the sound was indeed easily associated with a truck. For “stop sign” and “bench” it was difficult to find a good conceptual mapping. The sound for “stop sign” was the “beep” tone that some car navigation systems make to indicate a traffic sign. The sound that we used for “bench” is very similar to the one for “chair”. We selected them this way as the two concepts are similar, but this caused them to be frequently confused - see also Table 5.1 for the complete list of conceptual mappings used.

We logged the time for four out of five participants until they learned the correct object for each auditory icon. It took them on average 8.5 minutes (SD 2.15 minutes) to name correctly the classes for 17 auditory icons (we excluded “stop sign”, which was simply introduced by the test leader, as it is not intuitive). Explanations and comments from the test leader as well as the time for playing the sounds are also included in the guessing time. *The average learning time for all* 18 *objects for these* 4 *participants was* 3.2 *minutes.* Given the learning times and the rates for both intuitiveness and learnability, we infer that the sounds used were easy to learn and some of them even intuitive. It will have to be further analyzed whether the poor intuitiveness rates for some sounds are due to the conceptual mappings used or the sound design. The poorer intuitiveness averages (i.e. before the users knew the conceptual mappings) compared to the learnability averages afterwards indicate that knowing the conceptual mapping helps users to learn the sounds faster. This can be best seen for the "truck" class.

Comparison between auditory icons and spearcons. For spearcons, the average intuitiveness rates for all users and all objects (for P1 we only have the intuitiveness rates for 5 objects) was 2.5 out of a maximum of 3.0 (SD 0.5). This is much higher than for the auditory icons, which only had an average intuitiveness rate of 1.8 (SD 0.6). This could also be due to the limited (although not very small) set of classes. It should be further investigated if this scales well for larger objects datasets. All five users said, however, that they prefer auditory icons to spearcons. Participant P5 explains his preference: "*I think the others [auditory icons] are more feasible, because it's easy, you don't have to interpret their meaning, it's intuitive. You have to learn it first, but once the brain made the connection, it's no different than when I hear a car from behind.*" Thus, spearcons are in any case intuitive and fast, and hence suitable for sonifying objects, but another study is necessary to investigate whether they are better than auditory icons, which were preferred by our users. There could be a bias in our evaluation, as the spearcons were tested at the end, when the users were already accustomed to the auditory icons. Another study will thus have to be conducted, in which both modalities are tested in similar conditions.

Most useful objects. Table 5.3b shows the assessment of the participants when asked about the importance of objects during navigation. The *most useful objects selected by our participants were stairs, traffic lights, bicycles and motorcycles*, all crucial during critical situations. However, the ranking only shows the average usefulness. When analyzing single objects, for instance looking at the object "walls", it did not rank very high, but for one participant (P5) it was the most important object class, as it informed him about the existence of buildings, which in turn helped him to orientate.

When asked in what situation the users found objects useful, the answers differed. P2 for instance only wanted to know about stationary bicycles, while P3 only about moving ones. P4 could hear walls anyways and, thus, did not need to know about their location at all, while P3 only needs walls when it is dark outside, and P5 found them the most useful object class, as they provide him with information useful for orientation. This shows that the choice of objects is very personal and diverse, and depends on the users' navigation attitudes (Williams et al. 2013). Thus, it is essential that a system offers users the possibility to *choose* from a larger set of object classes.

When asked to name three other objects that they find relevant and were not part of the test set, the participants mentioned: pole (P2, P3, P5), garbage can (P4, P5), tree (P3, P4), water: brook, river, lake (P2), puddle (P1), fountain (P3), mailbox (P1), bus stop (P3), open trunk (P4) and truck's loading dock (P4). Two participants (P1,P3) wished the system could recognize text as well, for instance tram numbers, house numbers or text on traffic signs.

Assessment of the system in general and its interface. The participants appreciated the idea of the system in general. One of them said "*I like that you get a lot information from the environment, what* *happens left and right*" (P1), and another one added "*it tells me things that I don't perceive otherwise*" (P5).

The audio interface based on auditory icons and parameter mapping was rated very favorably. Two participants (P3, P4) said that they were positively surprised by the interface, as they expected speech to be used. Others also liked the panning (P1, P3, P4) and distance to loudness mapping (P3). Two other participants (P1, P2) praised the interface in general and said they were astonished that **one can learn the sounds so fast**. P5 remarked that "*sounds don't disturb much; one gets used to them quickly*". He also commented that "*it is similar to sight: one turns the head and perceives what's there - direct feedback - quite cool*". Thus, we conclude that the sonification chosen enabled the users to profit from the interface very fast. The comments strongly support the assumption that auditory icons combined with parameter mapping are suitable for object localization.

All of the participants named at least two objects that they found very useful. They appreciated that one can choose from so many object classes, and *configurability of the system was assessed paramount*. Each participant had his own set of preferred objects, according to the ranks and also comments. We also found that the *preferred distance for the object classes is different for each person* and may depend on the degree of impairment: P2 and P4, who are blind from birth, would mainly use the system short range (up to 10 m for most things), while P3 and P5, who have some minimal residual sight, found larger distances best. Regarding the angle at which objects should be identified, 4 out of 5 participants said that at least short range, the angle should be wider than the one of the white cane, if possible full range. P2 commented that for searching objects such as a park bench, the angle should be wider, while for obstacle avoidance the same angle as the span of the white cane is sufficient, but at a greater distance.

Users appreciated the inconspicuousness of the camera integrated into sunglasses. However, they suggested that the lenses should be exchangeable with transparent ones when needed, so as not to disturb their residual perception of light, which is helpful for navigation (P1, P2). Moreover, they also proposed that a camera should be wireless, and one of the users mentioned that the temple stem interferes with the frame of the bone conduction headphones. One participant (P1) wished he was able to change between a chest camera and one mounted on glasses. Two participants (P3, P5) proposed a headband similar to headlamp or only a light frame under the eyes instead of the full glasses (P5).

The participants were interested in further using the system to orient themselves (by hearing walls, traffic lights), to find things (bench, car/taxi, person) and to avoid obstacles (bicycle, car, person). According to P1, the interface can be improved such that dangerous dynamic objects are more prominent. The comments of the participants in general point out that orientation strategies and needs differ very much among people with visual impairments and, thus, supporting systems require a high adaptability, which corresponds with the findings of Williams et al. (2013).

Discussion and design implications. We investigated through an exploratory study how information about the environment can be conveyed to people with visual impairments when walking outdoors. We demonstrated the feasibility of using a camera-based system to localize multiple objects and to pass this information on to the user via sonification. We observed several issues regarding the camera, the creation of the sounds, the conceptual mappings of objects to sounds, dangerous objects and the interaction with the user interface, which we illustrate in detail in the following paragraphs.

Camera. The advantage of a head-mounted camera is that the camera follows the movement of the head, widening the angle of the surrounding. As a participant expressed it, “it is similar to sight: one turns the head and perceives what‘s there” (P5). Moreover, a camera worn on the body is better for people with visual impairments than one held in the hand, for two reasons: (1) it leaves one hand free, as the other one is often busy holding the white cane and (2) people with visual impairments have difficulties in focusing and holding the camera straight, so a camera attached to the body provides better images. To compensate for the disadvantages of the sunglasses-camera mentioned by our participants, we propose to use glasses with interchangeable lenses (transparent and toned) and integrate the earphones in the temples, using either directional or bone conducting sound.

Creation of auditory icons. We showed that the sounds can be memorized very easily and that on average, they are shorter than speech. Even more, when considering only the 5 most intuitive and easy to learn auditory icons (door, bicycle, dog, motorcycle, and bush), their average duration is 372 ms, while the corresponding speech lasts on average 627 ms - 69% longer (SD=70 ms). Or considering the sound for dog: the auditory icon was the shortest of the 18 sounds, namely 240 ms long, and had an intuitiveness of 2.4 and learnability of 2.8 (out of 3.0). This shows that one can find auditory icons that are very short and still intuitive, thus easy to remember. One participant, P4, expressed his preference for auditory icons as follows “I generally prefer sounds. Somehow you perceive them directly while you have to interpret speech” (meaning spearcons). The main challenge is to find sounds that preserve the meaning but are short enough, such that sufficient information about the environment is passed on the move.

Advanced users who listen to speech at rates of up to 500 words per minute (Edwards 1991) might profit more from using spearcons than auditory icons (the average speech length of the 18 object names in German, at a speech rate of 500 wpm is about 200 ms, while for auditory icons it was 538 ms). However, since not everyone uses the same speech rate, well-designed auditory icons offer the possibility to address all users.

Conceptual mappings of objects to sounds. The conceptual mappings of objects to sounds is a topic that has not yet been explored in detail. For instance, for bicycles, we chose the sound of a spinning wheel, where a flexible object is held in the spokes. The users of our evaluation found this sound very intuitive (average rate of 2.8 out of 3.0) and easy to use (average rate of 2.4 out of 3.0). In the literature, however, a bicycle bell is commonly used (Cabral and Remijn 2019). Moreover, the experts who pretested the prototype also made suggestions for different conceptual mappings than the ones we used for a couple of objects. This shows that **it is necessary to develop a better understanding of the conceptual mappings of objects to sounds that are most intuitive for people with visual impairments**.

Dangerous objects. Currently all objects are treated the same. If they occur in the scene, they are sonified and the user will be informed. However, the user must be able to recognize when objects can pose a danger, for instance, when they move fast towards the user. One possibility would be to track these objects and change the property of the sound or combine it with a certain warning sound.

Interaction with the user interface. Our study showed that the needs and capabilities of people with visual impairments are very diverse. The users exploited the options to switch on and off certain objects while walking (which was accomplished by the test leader during the experiment, at the users' wish). For instance, one of the participants dropped the object class “car” because there were too many parked cars along the way, thus this information was not useful. This shows that adaptation of the interface to the information need of the user is very important and has to be very simple. The user should have the option to make changes on the move without stopping. Moreover, an assistive system must contain a large pool of object classes from which the users can choose easily.

Sonification versus speech. While speech can convey almost any information, no matter how complex or long, auditory icons seem to be well suited for representing a large number of objects from the surroundings in a short time. Their advantages over speech are the following:

1. Auditory icons can be used in parallel with the speech output of pedestrian navigation apps. Thus, users get complex turn-by-turn navigation instructions in speech and an overview of objects in the environment as sonification. Since sonification is a different modality than speech, the cognitive load of the user should be lower than when using speech for both navigation and objects;
2. Auditory icons are short enough and can be shorter than speech in given scenarios, as shown by our analysis;
3. Auditory icons offer a direct perception of the object, unlike speech which must be interpreted first.

Limitations. The promising results of our initial work indicate many possibilities for future investigation. The most obvious one is the hardware: while the system is mobile, carrying a laptop in a backpack is unacceptable. Some users also complained about concomitant use of sunglasses and bone conducting headphones. One solution would be to use glasses that have integrated earphones.

The number of participants in our study could also be regarded as a limitation. A final system should be tested with more users, and with more diverse demographics. However, for a prototype, we consider that five users provide us with enough feedback to drive the development in a positive direction and inform future research.

A comparison between auditory icons and spearcons requires a subsequent test, where both sonification approaches are tested similarly with users - whether with the same users but in interchangeable order or with different users. The problem with our test is that spearcons were only tested at the end, when the users were already accustomed with the auditory icons, so our conclusion might be biased.

The think-aloud method that we used during the user study has several disadvantages: first of all, people may not say exactly what they are thinking (Shneiderman et al. 2018); secondly, describing one's thoughts may be difficult and "alter the process" (Shneiderman et al. 2018); finally, it is difficult for a person with visual impairments to talk while walking, since talking interferes with the user's awareness of the environment. While we cannot influence the first two factors, for the third one, we made sure that the users walk on routes that are well known to them. We also observed that they often stopped in order to make a longer comment. Since we did not record the times walked, this fact did not influence our study.

Conclusions. We presented a mobile prototype to support people with visual impairments in both micro and macro navigation in an outdoor environment. Our framework combining computer vision with sonification shows great promise in revealing environmental information otherwise inaccessible to people with visual impairments. Our exploratory study demonstrated that parametric auditory icons are well suited for conveying information about objects in the scene to the target users. We also showed that auditory icons can be learned very fast, despite their shortness. A major finding is that it is crucial for the system to be adaptable to the user's current needs regarding the object classes and the distance at which objects are sonified. We also found that objects relevant for the pedestrians' safety such as stairs, traffic lights, bicycles and motorcycles were considered very important. Moreover, auditory icons were found better than spearcons by all five participants in the given test setting, even though spearcons were more intuitive than the auditory icons. One important finding was also the fact that the conceptual mappings of the auditory icons need to be better investigated, which we do in the coming subsection (5.2.1). We showed in this chapter that sonification and especially auditory icons help people with visual impairments to better understand their environment.

Next, we show how we created, in a user-centered approach, a dataset of short auditory icons for several everyday objects for people with visual impairments. The dataset can be used in the development of assistive solutions for people with visual impairments in the travel context.

5.2. Creation and Evaluation of Short Auditory Icons for People with Visual Impairments

This section is partially based on the bachelor thesis of Lukas Fritsch (Fritsch 2020), which was jointly supervised by Angela Constantinescu and Dr. Karin Müller.

After we saw in the preliminary user study that short auditory icons have a great potential in travel applications for people with visual impairments, we decided to create a dataset of auditory icons that can be used in the development of such solutions. We did this in three steps: in a first questionnaire, we asked users from the target group about the sounds they associate with everyday objects (see Subsection 5.2.1). Then, in a user-centered approach, with the constant involvement of a blind person, we created a set of short auditory icons (see Subsection 5.2.2). Finally, we evaluated the resulted sounds in another online questionnaire with the target user group (see Subsection 5.2.3).

5.2.1. Assessment of Conceptual Mappings through Online User Study

As seen in the previous section, conceptual mappings still had to be further investigated. Thus, we started another user study, in which we asked people with visual impairments what sounds they associate with 42 everyday object categories. Our intention was to create auditory icons that are culture independent, or at least to see if there are cultural differences in the answers. Thus, we created the study online, in order to reach people from as many countries as possible.

The list of 42 objects whose conceptual mappings were to be evaluated contained both indoor and outdoor objects. It was obtained from the lists of the ATMaps project (Project ATMaps), namely the indoor ATMaps list which we had commented by experts - see Subsection 3.1.2 - and a list for outdoors - see Charitakis and Papadopoulos (2017) -, to which we added the list of outdoor objects that we evaluated in Section 4.1.3. While the ATMaps lists contain static objects, namely those that can be presented on a tactile map, in our evaluation we also had dynamic objects. The final list of 42 objects was obtained by clustering together, removing and restructuring the information from the three lists. For instance, "bus stop" and "metro station" were dropped, as they are already included in "public transportation station". We also excluded items that could not be associated with a point or area in space, such as "ground (walking area)", or that cannot be sonified, such as "street name", which can be conveyed intuitively through speech only. Finally, we only included the objects with the highest ratings to avoid overwhelming the participants in the user study. The final list contained thus the following 42 objects or object categories:

- Indoor:
    - Toilet or bathroom for men, women or people with disabilities
    - Elevator
    - Service point (reception, security office, service, information point, queue management ticket machine)
    - Room, room number and room type (office, conference room, laboratory, library, balcony)
    - Eating place (canteen, cafeteria or coffee shop, kitchen, vending machine)
    - Payment and cash withdrawal point (cashier's desk, ATM, cash machine)
    - Office electronics (equipment or machinery, computer, printer, photocopier)
    - Household appliance (refrigerator, television, washing machine)
    - Electrical installation device (socket, light switch, intercom)
    - Fire safety equipment (fire extinguisher, fire hose)
    - Water source (sink or washbasin, shower, water dispenser, water tap)
    - Heat source and hot items (radiator, heating panels, fireplace; heating burner; oven; hot water or anything that could be hot to burn)
    - Window (window, window shutters, rolling blinds, blinds)
    - Handrail for indoor stairs, railings
    - Furniture (sofa, couch, bed, wardrobe, closet, locker; hanging cabinet; cupboard; drawers or dresser; shelf; book case)
    - Table (table, desk, teacher's desk, tea table)
    - Glass element (glass partition, glass surface, wall mirror)
    - Indoor pillar
    - Projecting wall
    - Floor number (label near the stairs with tactile symbol for stairs and floor number in Braille)
- Outdoor:

  - Bicycle
  - Vehicle (motorcycle, car, truck, bus, train)
  - Train tracks
  - Crosswalk (traffic light pole, zebra crossing)
  - Stop sign (traffic sign)
  - Fence
  - Public transport station (train, metro, bus, taxi station)
  - Building
  - Pavement or sidewalk

- Indoor and outdoor:

  - Person
  - Dog, other animals
  - Door (simple, automatic, revolving, sliding, glass door, automatic bars, entrance or exit, emergency exit)
  - Stairs, steps, level change
  - Obstacle (hanging or overhead; on the ground; short obstacle)
  - Seating (chair, sitting bench, armchair)
  - Ramp, slope
  - Shop (clothing; electronics; food: supermarket, grocery shop, market place)
  - Dangerous point or area
  - Trash can (garbage can, recycle bin)
  - Tactile or haptic information (tactile paving; tactile signpost; haptic reference point or landmark; change of the surface or texture on the ground; carpet; floor condition)
  - Plants (bush, flower pot)
  - Smoking area

The study was conducted as an online SoSciSurvey questionnaire implemented on a local server at the Karlsruhe Institute of Technology (KIT). The survey was conducted between 18.06.2020 and 30.09.2020. Our goal was to have an international study with participants from as many countries as possible, and 2 to 3 answers from each country. We thus advertised the study with this goal in mind. We sent invitations mainly per e-mail to international contacts of the Study Center for Visually Impaired Students (short SZS, now ACCESS@KIT) and also personal contacts of the researchers involved. Later on (end of July) we also advertised the survey on the Facebook page of SZS and the personal Facebook page of a researcher involved. Altogether, the survey was sent to contacts from the following countries: Germany, Austria, Romania, Italy, Belgium, China, US. Additionally, the Facebook accounts used to advertise the study have contacts from the visually impaired community in the Czech Republic, UK, Greece, Republic of Moldova, Poland and Japan.

The survey was created in English only, but the respondents could also answer in German or Romanian, besides English. This was stated in the introduction of the survey. In a pretest, a blind person filled in the survey using a screen reader and assessed it as accessible. Another 3 sighted persons tested the survey for comprehensibility and general troubleshooting. All prospective participants could take part anonymously in the survey by accessing a link. Before starting the survey, the participants had to confirm that they read the privacy statement and also that they have not previously participated in the survey. All data was recorded and stored anonymously. At the end of the survey, participants could optionally leave their e-mail address for one or several of the following three reasons: (1) in order to take part in a lottery; (2) if they were interested in the results of the survey; and/or (3) if they wished to take part in further surveys. The e-mail addresses were stored separately from the survey data to ensure anonymity. As an incentive for participating in the user study, we offered two Amazon vouchers for 25€ each, to be randomly drawn among the participants.

The questionnaire was divided into 4 parts. The first three parts contained mandatory questions with free-text answers about the sounds that best describe the 42 everyday objects. The fourth part contained 12 socio-demographic questions about age, gender, home country, extent of visual and hearing impairments and questions about mobility, aids and use of technology. All socio-demographic questions were optional, except for the question asking the degree of visual impairment. This question was mandatory because we wanted to evaluate the answers from our target group only. The socio-demographic questions of the fourth part were simple-choice (8 questions), multiple-choice (1 question) and free text (3 questions). We placed the socio-demographic questions at the end hoping that participants would be more likely to fill in this sensitive data after they saw and filled in the main questions in the first parts. We thought that the main questions would convince the participants in the survey of the scientific scope of our study and would also motivate them to finish the questionnaire.

At the end of the survey we had 10 complete and 1 partial (filled in 20 from 42 questions in parts 1-3) responses from people with visual impairments (with less than 30% visual acuity). We only report these answers here, all others are not considered.

Participants' demographics. From the 11 participants with visual impairments, 5 came from Austria, 3 from Romania, 1 from Germany and 2 did not indicate their country of origin. Thus, there was a bias on European cultures, but at least both East and West Europe were represented. Nine participants had ages between 21 and 61 years old, with an average (AVG) of 38 and standard deviation of 15.6 years. The other two users did not state their age. Out of 11 participants, 8 were male, 2 female and 1 preferred not to answer this question. Regarding the extent of the visual impairment, 2 had low vision (5%-30% visual acuity), both of them from birth, and 9 were blind (<2% visual acuity), 7 of them from birth and 1 shortly afterwards, namely by the age of 6. There is an obvious bias towards people blind from birth which could be explained by the use of the auditory channel. People blind from birth compensate for the missing visual information by using the auditory channel extensively, while late blind people must first learn how to do that, which is more difficult with increasing age. Finally, people with low vision focus on using as much of their residual vision as possible, so they pay less attention to sounds. This might be the reason why mostly people blind from birth took our survey, as they are more interested in the topic and can also answer more easily the questions related to sounds.

Results of the conceptual mappings questionnaire. The free-text answers to the question: "what sounds best describe the following objects?" were translated from German and Romanian to English, where necessary. A thematic analysis after Braun and Clarke (2006) was performed for each of the 42 objects from parts 1-3. In part 1, the answers form 11 respondents were considered. In parts 2 and 3 we had 10 answers. Table 5.4 shows an example for the object "toilet or bathroom". The second

Themes	**#**	**Percentage**
Toilet flush	7	41.2%
No toilet flush - indiscreet, embarrassing	1	5.9%
Washing hands - sink - rushing water - mountain spring	6	35.3%
Text-to-speech: differentiated sex / people with disabilities / generic	2	11.8%
Hum of neon lights	1	5.9%

Table 5.4.: Example of thematic analysis outcome for the object "toilet or bathroom for men, women or people with disabilities". The second column lists the number of people who mentioned the theme.

Keyword	**Uses (Count)**
toilet	11
toilet flush	10
flush	10
water	4
washing hands	3
flushing	3
sound	3
washing	3
hands	3
flushing toilet	2

Table 5.5.: Example of keyword analysis outcome for the object "toilet or bathroom for men, women or people with disabilities".

column lists the number of respondents who mentioned the theme. In parallel, we also conducted a keyword analysis using a free online service [4]. Table 5.5 shows the first 10 results of the keyword analysis for the object "toilet or bathroom". However, we considered that a keyword analysis does not represent as good as the thematic analysis the meaning of the words. For instance, in the example above, for "toilet", there are more keyword counts for flush than in the thematic analysis because some users repeated the word in their answer. Also, one user said "**no** toilet flush", and this does not show up accordingly in the keyword analysis. Thus, we decided to give up keyword analyses and use the thematic analyses only.

The complete thematic analyses for all objects can be found in Appendix A.2.

For each of the 42 object categories, between 2 and 12 themes were created. However, many categories often have separate themes for different objects of the category. For example, 5 themes were created for the "eating place" category in general, and 3 themes each for "kitchen", 1 for "coffee-shop" and 3 for "vending machine" - see Table 5.6. In such cases, we created several auditory icons per category, to

---

[4] `https://seoscout.com/tools/keyword-analyzer`

Category	Themes	#	Percentage
Eating place in general	Clattering of dishes and cutlery	7	30%
Eating place in general	Glass	1	5%
Eating place in general	People, human voice rumble	3	13%
Eating place in general	A food processor/blender	1	4%
Eating place in general	A gong, as was once called for dinner	1	4%
Kitchen	Clattering of dishes	1	4%
Kitchen	Dishwasher, oven or refrigerator	1	4%
Kitchen	Frying and cooking	1	4%
Coffee shop	Cups rattling on coasters	1	4%
Vending machine	Money falling into the machine	1	4%
Vending machine	The rattle when the product is taken out or money	1	4%
Vending machine	The hum of a vending machine's cooling	1	4%

Table 5.6.: Thematic analysis outcome for the object "eating place (canteen, cafeteria or coffee shop, kitchen, vending machine)". Users suggested different themes for different objects in the category. The third column lists the number of users who mentioned the theme.

reflect the answers of the users. For each auditory icon, one or more of the most frequently mentioned themes were used.

When creating the sounds, we considered the themes suggested for all objects in all categories. Thus, when creating the auditory icon for one object, we tried to avoid the themes used for the creation of all other objects, so that every object has unique conceptual mappings and can thus not be easily confused with others. However, this was not always possible. For "room", for instance, the most prominent theme was "door", which makes the sounds for "room" and "door" be very similar. We thus recommend not using the auditory icons that are similar and can be confused (e.g. for "room" and "door") in the same application.

5.2.2. Creation of Short Auditory Icons Set for Everyday Objects

Requirements. The first step we took before starting with the creation of the auditory icons was to gather requirements from literature. We came up with the following list - also see Fritsch (2020):

- Sounds should have a duration of at least 400 ms in order to be interpreted correctly (Fabiani et al. 1996) and at most 1000-2000 ms, if user's immediate reaction is necessary (Cabral and Remijn 2019)
- Sounds used in public spaces or workplaces should cover at least the frequency range of 300-3000 Hz (Begault et al. 2012, Cabral and Remijn 2019)
- Iconic (or nomic) mappings of objects to sound should be used as much as possible, followed by metaphorical mappings and only scarcely symbolic ones. This ensures that the resulting auditory icons are as culture-independent as possible and can be used in an international context. Also,

try to represent the physical characteristics of the objects, for instance use wood knocking for wooden objects (Gaver 1986). In our case, this point should be considered when choosing among the mappings suggested by users in the conceptual mappings questionnaire (see previous subsection 5.2.1).

- In the creation of the auditory icons, multiple mappings and sound effects should be used in order to create more intuitive sounds. These should not necessarily correspond exactly to reality, but should trigger clear associations in the listener (Gaver 1989)
- Each auditory icon should be clearly identifiable; similarities should be avoided and contextual information should be added in order to differentiate between sounds that are otherwise very similar, such as heavy rain and the sound of hot fat in a frying pan (Hermann et al. 2011)
- The usage of more than 20-30 auditory icons in one application should be avoided (Frimalm et al. 2014).

Creation of the auditory icons. The auditory icons were created by a hobby musician and producer, student in sound and music production. The creation of the sounds was done iteratively and in a user-centered way, with the constant involvement of a blind user who is a hobby musician and has experience in sound production. Additionally, one expert in accessibility and assistive technologies was involved throughout the design and creation process to ensure that the needs of the user group are thoroughly addressed.

The auditory icons were created based on the thematic analysis, for each of the 42 object categories. Sometimes, several sounds were created for each object category, based on the answers from the questionnaire. For instance, for the “eating place” example, we created auditory icons for “cafeteria (eating place in general)”, “kitchen” and “vending machine”. For each auditory icon we created two versions: a shorter, **500 ms** one, and a longer, **1000 ms** version. The shorter version is useful for conveying as much information as possible in real time in an augmented reality scenario such as the one we described in Section 5.1, or for conveying additional information in interactive audio-tactile solutions like the one of Melfi et al. (2020). However, shortening the sounds so much can make it difficult to preserve their meaning. Especially since users are so diverse, it is very difficult to condense in such a short sound all the cues that make the meaning recognizable by every user. For this reason, we also provide the longer, 1000 ms versions, to be used in applications which are less time-critical. For auditory icons that only consist of a short sound (e.g. a beat) and therefore only last 500 ms or less, the 1000 ms version was not created anymore. This was the case for “dog” and “indoor pillar”.

For the creation of the auditory icons only self-recorded sounds or sounds from freely available datasets with CC0 licenses were used. These composing sounds match the themes from the thematic analysis and were first edited individually. By equalizing (processing the volume of different frequencies or frequency ranges), unnecessary or disturbing frequencies were lowered and important frequencies were emphasized. By unnecessary frequencies are meant frequencies that do not contribute to the actual intended sound and possibly unnecessarily "pollute" the final signal. For example, if the focus is on a high-pitched buzzing, but there are low frequencies in the sound, those low frequencies will be lowered or cut out altogether, as long as the actual sound is not affected. Disturbing frequencies are frequencies that are too loud in the original sound and are either unpleasant to hear or draw too much attention to themselves (stand out). In the same example, if the high-pitched buzz is a bit too loud, or too shrill in the upper frequencies, these will be lowered. Finally, by important frequencies are meant frequencies that define the actual sound, such as the middle frequencies in a phone ring, the lows in a hum, the highs in a swish or clatter. These were emphasized, if necessary. The volume and length of the individual sounds were adjusted. After the individual editing, the sounds corresponding to single themes were merged and any possible collisions such as overlaps of frequencies from several sounds, which result in disturbing noises or make the original sounds less identifiable, were removed. A reverb effect was applied to some sounds to better illustrate the spatial characteristics around the sonified object, but only when the reverb attributes were specifically mentioned in the survey responses (Fritsch 2020).

Feedback from the blind expert. After all auditory icons were created, they were reviewed in two feedback sessions by the blind user who had many years of experience in music and sound production, and whom we refer to as *the blind expert*. We refer to the student who created the sounds as the *sounds creator*. The goal of the feedback session was to assess the quality and intuitiveness of the sounds used to create the auditory icons and whether they kept their comprehensibility after processing. Since the feedback was done in the middle of the Covid pandemic, it took place online. Due to the extremely short and many sounds and the use of a screen reader by the blind expert, it was impossible for the latter to listen to all the sounds independently, as the screenreader talked while the sounds got played. Thus, the following procedure was used:

- The sounds creator sent the auditory icons in a zip file to the blind expert
- Both installed a remote control software (Teamviewer) and the sounds creator accessed remotely the blind expert's computer
- The sounds creator played the auditory icons on the blind expert's computer while the screen reader was deactivated
- While hearing the sounds, the blind expert and the sounds creator talked on the phone, as talking through the remote software did not work while the sounds were playing. It would be more convenient if one could talk to each other through the remote software while playing the sounds at the same time
- After hearing each sound, the blind expert said what object or situation he associated with it. The sounds creator then explained which object the sound should represent and the blind expert gave feedback on it whether it is well recognizable, misleading, or how it could be improved. For instance, for the object "bicycle", the blind expert suggested to play first the chain sound, while emphasizing the high frequencies more, and then the bell sound, or use a completely new sound or mapping. Sometimes, the blind expert made suggestions on how to record sounds, or what alternative sounds to use for the given conceptual mapping. For "ramp", for instance, he suggested using the sound of a marble rolling down a board or a skateboard's noise.

As a result of the blind experts' feedback, 12 auditory icons were modified, and a second feedback round was conducted, this time for these 12 objects only. After the second feedback round, the blind expert suggested one alternative sound for one auditory icon, and also increasing the volume for 3 further auditory icons. In the meanwhile, one expert in accessibility and assistive technologies with expertise in sound interfaces for people with visual impairments also gave feedback to the auditory icons. The last suggestions made by the experts were implemented, resulting in our final dataset of auditory icons for everyday objects to be used by people with visual impairments in the travel context.

5.2.3. Evaluation of the Auditory Icons Dataset

The goal of this user study was to assess the comprehensibility (intuitiveness) and attractiveness of the auditory icons we created, while at the same time identifying conflicting mappings. For the evaluation, we adapted common evaluation techniques used for visual icons, such as Nielsen (1995), UsabilityHub, while also taking into account the recommendations specific for auditory icons, such as: "Evaluate the identifiability of the auditory cues using free-form answers" (Hermann et al. 2011). In order to solve the problem of too similar sounds for different concepts within the same category (e.g. office table vs eating table, sofa vs bench), and also to reduce the number of sounds to be evaluated, we decided to choose one sound per category only to be evaluated. We did this in a pre-study with sighted people.

Pre-evaluation: choosing one sound per category. In the pre-evaluation, our goals were: (1) to find the sound most representative for the given category; and (2) to evaluate identifiability within the category only, so that all sounds, including the ones left out in the subsequent main evaluation, are assessed in the end. There were 9 categories containing more than 1 auditory icon each (namely between 2 and 5 auditory icons per category), adding up to a total of 24 auditory icons. The participants to the pre-evaluation had to answer two questions: (1) After hearing all auditory icons from one category, order the sounds descendingly, starting with the ones that best describe the category. (2) Assign each sound to an object within the category. For instance, in the category "eating place", assign the corresponding auditory icons to "cafeteria" and "kitchen". Thus, if auditory icons are not correctly assigned in this step, they are either not intuitive for the object they should represent, or they have conflicting mappings with other objects in the same category. Finally, users had the possibility after each question to make comments in a free-text field. We expected this to happen when a sound was not intuitive or did not meet the user's expectations.

We received 18 answers to the pre-evaluation, 12 of whom complete and 6 partial. Based on these answers, we could choose one object most representative for each category. Additionally, based on the second question and on the user's comments, we assessed the intuitiveness and conflicting mappings for the auditory icons involved. In their comments, for instance, users often wrote what they associate with certain sounds. For instance, in the "table" category, one user said: "Tone C reminds more of a door because of the 3-times knocking". Thus, there are conflicting mappings between "table" and "door". Finally, the comments also helped us get suggestions for improvement, such as: "I would prefer the sound of a glass being placed on a table. The second sound in the recordings seems irritating to me."

Main evaluation of the auditory icons online. After reducing the number of sounds to be evaluated in the previous step, we further split the category "electrical installation" into "electrical installation (socket)” and “electrical installation (switch)”, as it is important for people with visual impairments to differentiate between the two. Additionally, we removed another 3 auditory icons, namely the least important ones: “projecting wall”, “fence” and “public transport station”. Our goal was to obtain a list of only 40 auditory icons (for indoor and outdoor objects) to be evaluated, as we did not want to overload the users too much.

We implemented our questionnaire again in SoSciSurvey running on KIT servers. The questionnaire had a German and an English version. The survey was primarily aimed at blind people. We advertised the survey internationally per e-mail to contacts of the ACCESS@KIT center. Besides contacting propagators such as heads of visually impaired organizations, we also wrote personal e-mails to addressees in our contacts list who previously agreed to be reached for user studies. The questionnaire was pre-tested with 3 sighted and 1 blind person for: accessibility, comprehensibility and general troubleshooting. Before starting the survey, the participants had to confirm that they read the privacy statement and also that they had not previously participated in the survey. At the end of the survey, participants could optionally leave their e-mail address for taking part in a lottery for winning 1 of 5 vouchers of 25 Euros each, or if they were interested in the results of the survey.

The study was divided into three parts: (1) demographics, then (2) the second, most important part, containing questions about the auditory icons, and (3) a few general questions in the end asking among others about the type of audio output device and screen reader used (if any). In the second part, for each object, the auditory icon was played and the users were asked to answer questions about: (1) how well they liked the sound (Likert scale, 1 through 7); (2) if there is anything that bothers them about the sound; (3) choose from a drop-down list of five items, the fifth being a free-text "other", what term / object best matches with the sound. We chose in the drop-down list four objects from the 40 that represented concepts most similar to the auditory icon under evaluation or had the most similar conceptual mappings; (4) leave optional comments. The question about attractiveness (how well they liked the sound) was placed up front on purpose. If the users were first asked which object they associate with this sound, when asked afterwards if they liked the sound they would have included in their answer to the latter question how well the answers to the first question matched their expectations - which happened during the pre-test. The attractiveness question, however, only asks how pleasant the user finds the sound, and not how well it matches the object it represents. The auditory icons contained 1000 ms before and 700 ms after of silence, to allow the screen reader (if any) to speak without interfering with the sound playing.

Due to the still quite large number of auditory icons to be evaluated, the survey ended up being quite lengthy. Thus, when creating it, we took a gamification approach, in order to encourage users to keep on filling in. After each question asking the meaning of a sound, on the next page, we gave the correct answer, points were counted, and the user could play again the sound and optionally make comments. Since we still expected at least some users to fill in the questionnaire only partially, we randomized the order of the auditory icons, aiming in the end to we have a similar number of answers for each auditory icon, despite several partial answers. For this reason, we also placed the demographics section in the beginning this time, because at least the question about the visual impairment was very important to us. In the end, we had 16 answers from blind users - 11 complete and five partial answers.

Participants' demographics. From the 16 blind participants, 14 came from Germany, 1 from Austria and 1 from the UK - thus, there is a bias towards the German culture. They were between 18 and 87 years old, with an average (AVG) of 39.4 and standard deviation (SD) of 17.4 years. Eight participants were men, 7 women and 1 stated to be diverse.

Results of the online evaluation of auditory icons. The comprehensibility of the sounds was tested by playing the sounds and asking users what they represent. The auditory icons were played (listened to) by one user 2.2 times on average, with almost a third of all sounds being played only once. On average, participants answered correctly in **71.1%** of cases for all auditory icons. The *incorrect* answers often included:

- super-categories, e.g. when users answered “furniture” instead of “table”;
- the conceptual mapping used - for instance, users answered “closing door” instead of “building”, while the auditory icon for building includes indeed the sound of a closing metallic door;
- related concepts, such as “fire safety” and “dangerous point or area” or “crosswalk” and “vehicle”.

Still, all these answers were counted as incorrect, even though they relate to the object under question. They show us possible conflicting mappings.

Seven objects were correctly identified by all participants - see Table 5.7a. By far the most problematic were “indoor pillar” and “room”. Indoor pillar is very difficult to conceptualize and sonify, as it does not make any sound by itself, nor can it be easily associated with something that makes a sound. Besides, it can consist of different materials, most often, however, of concrete, which is again difficult to sonify (as opposed to wood, for instance). However, even so, the auditory icon for indoor pillar was most often mistaken for an “obstacle” (5 times), which is as close as it gets to an indoor pillar, as the main reason why a blind person would like to know about an indoor pillar is because it can constitute an obstacle that can be hit. Three users identified it correctly, two users associated the sound with a “seating” and one with a “stop sign”. In the case of “room”, it was mistaken 7 times for a “door” and twice with “building”, out of 12 answers. Only 3 users identified it correctly as “room”. It must be said, however, that a door was the main conceptual mapping used for “room” - and for “building” as well, making these 3 concepts most prone to conflicting mappings.

The average attractiveness for all auditory icons and all users was 4.6 with a standard deviation of 0.6, on a scale from 1 - not at all to 7 - very attractive. Thus, the attractiveness is above the average of 4.0, going towards 5.0 - somewhat attractive. However, the comments made by users suggest that this question was not always answered as intended - i.e., it does not reflect entirely whether the users find the sounds pleasant and nothing more. Although the question about attractiveness was asked first in the evaluation, right after playing the sound, users still coupled the attractiveness rating with how easy it was for them to find a meaning to the sound. This can be seen in the answer to the question: “Is there anything about the sound that bothers you?” which was asked right after the attractiveness rating. Users commented about the sounds that they are: “not intuitive” (P4), “baffling”, “meaningless” (P5), “ambiguous” (P7, P13), “difficult to assign” (P8, P14). Besides, the attractiveness scores are moderately correlated to the comprehensibility scores - the correlation coefficient being 0.43, which is a moderate positive correlation. Still, some of the comments reflected indeed the problem of sound

Object	% correct
Bicycle	100
Closing Door	100
Dog	100
Handrail	100
Household Appliance	100
Service Point	100
Vehicle	100
Eating Place	91.7
Glass	90.9
Smoking	90.9
Stairs	90.9
Toilet	90.9
Danger	84.6
Electrical Installation: Switch	84.6
Office Electronics	83.3
Water Source	81.8
Payment	76.9
Trash	76.9
Person	72.7
Tracks	72.7
Elevator	66.7
Fire Safety	66.7
Plants	66.7
Shop	66.7
Table	66.7
Electrical Installation: Socket	64.3
Obstacle	63.6
Seating	63.6
Stop Sign	63.6
Window	63.6
Floor Number	61.5
Ramp, Slope	54.5
Tactile Information / Surface	54.5
Furniture	36.4
Heat Source	36.4
Pavement	36.4
Building	35.7
Crosswalk	35.7
Indoor Pillar	27.3
Room	25

(a) Comprehensibility of the auditory icons, as average percentage of correctly chosen answer, sorted descendingly.

Object	Attractiveness
Dog	6.8
Closing Door	5.6
Room	5.5
Bicycle	5.5
Glass breaking	5.4
Obstacle	5.4
Tracks	5.2
Office Electronics	5.1
Ramp, Slope	5.1
Handrail	5.0
Service Point	5.0
Payment	5.0
Elevator	5.0
Household Appliance	4.9
Trash	4.9
Danger	4.9
Pavement	4.9
Vehicle	4.9
Eating Place	4.9
Table	4.7
Toilet	4.7
Fire Safety	4.7
Smoking	4.6
Seating	4.5
Furniture	4.5
Floor Number	4.5
Stop Sign	4.3
Electrical Installation: Switch	4.3
Person	4.3
Crosswalk	4.2
Water Source	4.2
Tactile Information / Surface	4.2
Shop	4.1
Building	3.9
Plants	3.9
Heat Source	3.8
Stairs	3.8
Electrical Installation: Socket	3.8
Indoor Pillar	3.6
Window	3.4

(b) Average attractiveness for all 16 users (from 1=I did not like the sound at all, to 7=I liked the sound very much. Note that 4 is in the middle, i.e. neutral) sorted descendingly.

Table 5.7.: Comprehensibility and attractiveness ratings of the auditory icons evaluated, with averages for all 16 users.

aesthetics: “sounds a bit artificial”, “too comic-like”, “sounds rather interesting” (P3); “is very shrill” (P6); “siren a bit loud” (P7); “sounds like violence”, “sounds nice”, “a very fine whistling sound”, “could have better quality” (P9); “too quiet” (P13). Often, users commented that the sounds are too short and thus difficult to interpret. While this is true, it was also the purpose of the study to evaluate how intuitive such very short sounds can be.

5.2.4. Discussion

Conducting research for people with visual impairments while involving the users throughout the development process can be a demanding task. We present here some of the problems that arose when creating very short auditory icons using feedback from users and experts, and propose solutions.

Few participants in user studies. When creating studies with people from a restricted user group, such as people with visual impairments, you do not have as many users as when doing research for the general public. Moreover, sometimes, you only need answers from blind people, which further restricts your options. Finally, if you conduct studies online, some people, such as many older people, are left out. In the case of our conceptual mappings questionnaire, for instance, its difficult nature together with its online format certainly discouraged or left out many users. Other issues that emerge when trying to acquire participants are data protection regulations, which differ from country to country, and lack of knowledge about the structures in other countries, such as how are organizations for your target group built, how can you contact them, etc. All this makes it difficult to obtain answers. Possible solutions to this problem are:

- Build up national and international contacts, especially propagators such as heads of organizations for your target group; maintain lists with possible participants and ask after each user study if participants agree being contacted for further studies;
- Build a trust relationship with your participants by always being fair, polite and sincere;
- Offer incentives. Besides financial incentives, users often participate out of interest for the topic under study. Thus, provide the results of the study to the participants who left their e-mail addresses for this purpose, and say a few words about how their involvement informs future developments. Do this in a timely manner;
- Contact prospective participants in personalized e-mails;
- Divide too long questionnaires and present different users with different parts of the questionnaire; or, randomize the order of the questions such that partial answers can also be considered, the way we did for the evaluation of the auditory icons;
- Carry out phone interviews instead of online questionnaires - or both;
- For interviews or on-site studies: conduct several studies about different or similar topics in one single session, at one location, if possible. Gather as many scientists in your department, city or area and organize it together. This way, it is more worthwhile and rewarding for participants to travel from further away to participate in several studies.

Language / communication issues. When conducting international studies, communication can be a great barrier. Especially in our case, for the conceptual mappings questionnaire, making sure that users understand what we need from them, and understanding what they were trying to tell us, can be very difficult. In our case, we asked the questions in English, but most of the participants were non-native in English, thus, they might not have understood everything correctly and completely. Moreover, the researchers who interpret the answers might also face problems due to: the usage of a non-native language; typos or wrongly used words; cultural and personal differences; use of thematic analysis where answers must be clustered into themes, but clusters might emerge that do not represent the most important aspects noted by the participants. All these can lead to false understanding and interpretation of the answers. Possible solutions to these issues include:

- Use professional translation services to translate both the questionnaire and the answers afterwards. Make sure that the translators understand well the purpose of the user study;
- Carry out interviews instead of online questionnaires, where the users can be asked questions.

Too many and too difficult studies. In order to avoid overwhelming users with long and difficult questionnaires, one could first create the auditory icons according to the requirements formulated in the beginning of this chapter, using as much as possible iconic mappings, and test with 5 users, as recommended by Nielsen (Nie 2019). If necessary, create a conceptual mappings questionnaire only for the sounds that come out as problematic in the test. This should reduce the number of objects to be evaluated, making a conceptual mappings study easier.

Creating the auditory icons. Problems in creating the sounds from the conceptual mappings that came out in the study include:

- Finding or recording appropriate, qualitative and intuitive sounds that have a suitable license (cc0 or similar);
- By shortening the sounds so much, especially in the case of the 500 ms long ones, they might become unintuitive;
- Combining several sounds into such a short auditory icon, as suggested by literature and by the participants in our user study can prove difficult;
- Several objects might share conceptual mappings (e.g. door, room, closet all have the creaking door as a main conceptual mapping), and can thus be confused with each other.

Possible solutions to these problems include:

- For the problem of finding appropriate sounds, we do not have a good solution, but we do encourage open data projects. Alternatively, record your own sounds - several for each object -, using professional equipment;
- If one sound becomes unintuitive after shortening, use another sound or try another conceptual mapping;
- For the 500 ms sounds, do not combine more than 2 conceptual mappings;

- Do not use the same conceptual mappings for several sounds, but find different conceptual mappings instead;
- If using the same conceptual mapping for different objects cannot be avoided, do not use sounds with conflicting mappings in the same application.

The successful use of the auditory icons depends, among others, on the context of use and users' proficiency with the system, i.e. their knowledge of the context of use and the objects that are sonified. To exemplify this: in the online evaluation of the auditory icons, when hearing the sound for "seating", 7 participants out of 11 identified it correctly among the possible answers, while 4 participants said it was "furniture" - which is not completely wrong, since seating is a type of furniture. However, we used a different auditory icon to differentiate between furniture in general (to be seen more like an obstacle) and seating, which can be something one searches for. Thus, in this case, the answer "furniture" was considered wrong. Another example: in the case of "vehicle", all participants chose the correct answer from the list. However, the auditory icon represented actually the engine of a motorcycle - which had been chosen in the pre-evaluation to represent the entire "vehicle" category. If "motorcycle" would have been among the possible answers, probably the answers would have been divided, similar to the "seating" example above. Thus, a good rule of thumb is to *avoid using both objects and their categories in the same application. If this cannot be avoided, make sure the users know the difference*.

Correlation between evaluation of attractiveness and comprehensibility. One solution to the problem that users rate a sound as being more attractive if they easily find a meaning to it, could be to test attractiveness separately from comprehensibility, identifiability, learnability or intuitiveness, that is, in a different user study.

5.2.5. Conclusion

We created, in a user-centered way, with the constant involvement of experts and end-users during development and evaluation, a set of very short auditory icons for 40 everyday objects[5]. In a first online user study, we asked 11 people with visual impairments about the sounds they associate with a set of everyday objects - these are the conceptual mappings of the auditory icons. Based on these answers, we created the sounds with the involvement of a blind hobby sound engineer. The sounds have two versions: a 500 ms and a 1000 ms one. In an online study with 16 blind users we evaluated the 500 ms sounds. The results showed that in more than 70% of cases, users associated the correct object with the sound, and in many other cases they associated related concepts. The sounds were previously unknown to the users and they were played on average only 2.2 times during the study. Thus, after hearing the very short auditory icons only briefly, users could choose the correct association in more than 70% of the cases, which is an encouraging result. Many sounds were found to be pleasant, while in few cases suggestions for improvements were made by the participants.

---

[5] The sounds will be made publicly available under an open license after their scientific publication in a conference or journal.

Our short auditory icons can be used in various scenarios for people with visual impairments, from mixed reality travel solutions to audio-tactile maps (Melfi et al. 2020).

5.3. Comparison Between Speech and Non-Speech Sonification

In Chapter 4, we have seen how speech can be used in both outdoor and indoor contexts to support people with visual impairments during travel. In the first section of this chapter, Section 5.1, users suggested that sonification might be preferred over speech in given scenarios. The goal of this section is to compare speech and sonification in terms of usability by assessing effectiveness, efficiency and user satisfaction (Brock et al. 2015).

Since sonification is not a widespread topic, research projects comparing non-speech sonification and speech are very scarce. One of them is the work of Fiannaca et al. (2014), who compared speech and non-speech sonification with 8 blind people in the context of crossing a large open space. The interface was thus conveying information about a fixed point on the other side of the open space. They found that the particular sonification chosen resulted in less efficient navigation than speech with respect to time, although the difference in veering between the two methods was not significant. During the sonification-based navigation, users had to stop more often to correct their orientation, while when using speech, users veered on the go. This might explain why users also expressed a preference for speech, although the difference was again not significant. Some users found the pitch mapping to distance from the sonification to be less obtrusive than the speech-based distance updates. The authors thus suggest using a combination of speech and sonification for best results.

In a similar context, namely guiding a blind person, but this time during road crossing, Mascetti et al. (2016) compared speech with 2 different non-speech sonification techniques. They found that none of the 3 modes outperforms the others, and the preference of the users is divided among the three techniques. They found that both speech and non-speech sonification have their advantages and disadvantages: “higher effort is necessary for decoding the sonified instructions if compared to the speech instructions [...]; with speech messages it is harder to hear the sound of the environment [...]; sonified messages convey information about the ‘quantity’ of the expected movement” (Mascetti et al. 2016).

We conducted an evaluation inside a building, comparing speech in form of simplified egocentric descriptions with sonification in form of descriptions based on very short auditory icons combined with parameter mapping. The idea was for the users to stop and trigger a description of the environment at that location, that is, get information about objects around him within a given radius (which we set to about 25 meters for our test). It is the same use case as for the MapEar prototype which implements the egocentric textual descriptions - see Subsection 4.1.2.

The evaluation was a sort of Wizzard-of-Oz study, in which we simulated the functionality of a system by recording instructions at fixed locations inside the building. Thus, the users could not call the descriptions at any position in the building, like in a real scenario, but had to stand at fixed positions in order to hear the instructions, which were triggered by the test leader. It was necessary for the comparison between speech and sonification that the same instructions are used for both modalities.

Figure 5.3.: Evaluation setup for comparing speech versus sonification: blind user standing at fixed position inside the test building, listening to the descriptions. On the floor, the position and the looking direction are marked.

The speech instructions included the object name, its distance and orientation with respect to the user (i.e. with respect to the fixed, preset locations inside the building). The study was conducted in German, but we provide the English translations for comprehension. Below is an example of a speech instruction:

> German: Toilette auf 12 Uhr in 18 Metern.
> English: Toilets at 12 o'clock in 18 meters.

Note that the object does not include details, thus this is a simplified version of the indoor grammar variant Object-Direction-Distance, as defined in 4.4. This was necessary, as it had to be equivalent to the sonification-based instructions, which use auditory icons for the object name (category) and parameter mapping for distance and direction, as defined in Section 5.1.

For the sonification-based descriptions of the environment, we needed to convey objects 360° around the user. This was a problem, as the bone conducting headphones cannot convey 3D sound very well - the user would not be able to differentiate between front and back. We thus organized a small expert round-table to clarify this issue.

Expert round-table to discuss 3D sonification. Three experts participated in the round-table: a blind hobby musician with experience in sound production and expert in accessibility; a hobby musician and producer, student in sound and music production; and an expert in human-computer interaction and accessibility with experience in sound interfaces. Because of the COVID-19 pandemic, the round-table took place online. The experts first discussed the usage of various headphones: bone conducting and headphones featuring directional sound do not convey 3D sound well. Especially the sounds behind the user are difficult to differentiate. AirPods Pro can convey 3D sound, but are closed headphones that cannot be used by blind people on the move, as they mask environmental sounds. They might be acceptable for virtual reality or simulation scenarios, but not for our context. A general problem with 3D sound is that general head-related transfer functions (HRTF) do not work equally well for everyone, while individual HRTF (that encompass the exact characteristics of a user's head needed to convey 3D sound, such as distance between the ears, form of the ear, etc.) are difficult to obtain and require expensive equipment. Finally, 3D sound cannot be used by blind people on the move, as it often induces a loss of balance and nausea, in the same way that 3D glasses produce in sighted people. Thus, the experts decided not to use 3D sound in this context. What was needed was to convey the x-axis displacement, which can be easily achieved with panning, and front-back differentiation. Y-displacement is not necessarily needed, as it is assumed that everything is at street level. The y-displacement for most objects can be deduced based on the object class. For instance, traffic lights are higher, trash bins or dogs are lower. For differentiating between front and back, several methods were proposed and discussed by the experts:

- Muffle the sounds behind. The disadvantage is that the muffled sounds might become unrecognizable. Besides, it might not be immediately clear which object is muffled and which one is not.
- In front use higher frequency, behind lower, similarly to the Doppler effect. The same problems may arise as for the previous proposal.
- Play an extra sound behind, for instance, a white noise. If playing the sound at the same time with the auditory icons, it could mask the latter. If playing before each sound, it makes the sonification unnecessarily longer.
- For the sounds behind the user, use additional vibration. This is only possible if the hardware (bone conducting headphones) had a vibrating motor that can be triggered at the same time while the sound is playing. One suggestion was to use the vibration of a smartphone, but that would be confusing, besides requiring the user to hold the phone.
- The user presses buttons for “front” and “back”. However, this requires unnecessarily more of the user's attention and handling.
- Finally, the experts decided that the best solution for this context (i.e. user stops, presses a button and hears a sonification-based description of the environment) is to use speech for announcing “front” (German: “vorne”), play the objects from 9 to 3 o'clock, then announce using TTS “back” (German: “hinten”) and play the objects from 3 to 9 o'clock.

Once the problem of sonifying the objects behind the user was resolved and the sonifications were recorded, the preparations for the study continued. The study consisted of two parts:

1. Learning the auditory icons used during the evaluation, 8 in total, each 500 ms long, chosen from the dataset described in the previous Section (5.2).
2. Main evaluation phase: walking 2 routes, about 50 m each, once with speech, once with sonification, starting in interchangeable order for each participant.

In the learning phase, each of the 8 auditory icons was played and the user had to guess what object it stands for. Then, the test leader explained the sound - what object it represents, and what conceptual mappings were used, that is, what can be heard in the auditory icon. Then, the sound was played once more before moving on to the next object. The 8 objects were: “elevator”, “entrance”, “corridor”, “wardrobe (checkroom)”, “room”, “toilets”, “stairs”, “water source”. As can be seen, 3 of the 8 objects did not appear in our dataset of 42 auditory icons. We decided to use similar sounds from our dataset, and see if they were as intuitive as the others, that is, if our sounds can be used in a flexible way, depending on the context. Thus, we chose the “building” sound for “entrance”, as it features a metallic door, typical for entrances. For “corridor”, we used the sound for “pavement”, which features high heels on a hard surface such as concrete or stone. Finally, for the “wardrobe (checkroom)” the sound for “locker” was used, which features the opening door of a wardrobe and sounds of clothes hangers.

In the main evaluation phase, each of the 2 routes had 5 points (locations), including start and stop, thus 5 descriptions. Each description had between 8 and 12 instructions, with a total of 49 and 52 instructions per route respectively. Before the main phase started, one demo for each route type was played, to allow the users to get used to the instructions. At two points on each route, namely at start and 3rd point (in the middle of the route), participants received a task after they have heard the description for that point. Examples of tasks are: “Please go to the restrooms and touch one of the doors”, “Please go to the elevator and call it”, “Please go to the farthest stairs and touch the handrail”. The task was considered successful if the user correctly carried out the action for the right object: “call the elevator”, “touch the handrail”, etc.

Participants’ demographics. Five people with visual impairments participated in the evaluation: 3 blind from birth, 1 late blind (onset of blindness in adulthood) and 1 severely visually impaired (visual acuity less than 5%, onset in later childhood). There were 4 male participants 1 female. The participants were between 18 and 60 years old with an average of 34.2 years. Three participants did not know the building in which the study took place, while two knew it moderately - but did not know the routes that were used during the evaluation.

Results of the evaluation. We report here the results of the evaluation, starting with the learning of the auditory icons. Then, in the main evaluation phase, we compared speech with sonification in terms of usability, namely: efficiency, effectiveness and user satisfaction. To test efficiency, we compared task completion times and the cognitive load (how much cognitive resources were spent for the task). For effectiveness, we used the success of task completion. For assessing user satisfaction, we used the Net Promoter Score (NPS), which is shorter than the established system usability scale (SUS), but, according to some, comparable to it (Brooke 2013).

Learning of the auditory icons. We played all 8 auditory icons to the users and asked them to guess what it represents. This time we did not read the list of objects to the users in advance, as we did in the preliminary study of the outdoor object sonification, as this time the list was too small. Instead, we told the participants that they should expect objects or things that can be found in a building. Table 5.8 lists the answers of the participants. Out of 40 cases (8 objects, 5 participants), in only 11 cases the object was correctly identified. In another 9 cases, either a related concept was named, or the sound used to create the auditory icon. In 20 cases, however, that is 50% of all cases, the users did not guess the meaning of the sound. The sounds were played only once or twice, and only in 3 cases they were played 3 times. Twice, the participants asked to raise the volume. The least identified object was

Object	Correct	Partially Correct	False
Elevator	4	0	1 (Ventilation)
Entrance	0	2 (2x Door)	3 (2x Playroom, Something sharp)
Corridor	1	2 (Steps, Walking over something)	2 (Stairs)
Wardrobe	0	1 (Closet door)	4 (Something hinged, Trash can)
Room	0	4 (4x Door)	1
Toilets	3	0	2
Stairs	0	0	5 (Trash can lid, Something rattling, Something with a door)
Water source	3 (Sink/water, Faucet, Shower/water)	0	2
Total	**11**	**9**	**20**

Table 5.8.: Number of times that participants associated an object with an auditory icon: correctly, partially correctly or falsely. Partially correct means that a related concept or the conceptual mapping used for the creation of the sound was named. The answers of the users, for the partially correct and false answers are in parantheses. Sometimes, users could not associate anything with the sound, so there are fewer answers in the "false" column than false counts.

“stairs”, which was falsely or not at all identified by all 5 users. In the online evaluation, however, the same sound was correctly identified in 10 out of 11 cases. One possible reason for this discrepancy could be the type of question - in the online study, it was a simple choice, where users had only few objects to choose from, while here, it was a free text answer with no hints whatsoever. Another reason for the difference could be in the use of the bone conducting headphones in a rather noisy environment in this study, as opposed to the own, regular headphones used in the online study by most participants. Bone conducting headphones convey the sound in a poorer quality than regular headphones, which could have made the association with the correct object more difficult. This points to the users’ need to learn the sounds. As could be seen in the preliminary study, after a short learning phase the users could use the sounds with high confidence. The three objects for which we did not have auditory icons, but we used similar ones, had a lower comprehensibility rate than the average. Wardrobe in particular was partially correctly identified only once, and 4 times it was not guessed. While this is, indeed, a difficult object, it also points to the fact that using the auditory icons for the purpose they were created is important for their efficient use.

Efficiency. In order to assess the efficiency of the instructions, we recorded the time needed to complete all tasks, as well as the cognitive load of the participants. The total time needed to complete the tasks was slightly better for sonification (42:14 minutes) than for speech (45:15 minutes), but with no statistically significant difference. However, for route one, on which the sonification was played more often (three times versus twice), the total task time was greater than for route two. Thus, if the routes would have been identical, the difference in task times between sonification and speech might have been greater. The cognitive load measured with the raw NASA-TLX was very similar for both modalities: 53% for speech and 54% for sonification. We thus conclude that in terms of both time and cognitive load, sonification is as efficient as speech. Given more learning time for the auditory icons, we expect the sonification to be more efficient in terms of cognitive load as well, but this is an assumption that will have to be verified in a long-term evaluation.

Effectiveness. To compare effectiveness, we counted the success in completing the given tasks. For the speech instructions, five out of 10 tasks have been completed successfully by all participants, while for sonification, seven tasks out of 10. As sonification was more often played on the first route (three times versus twice), and speech the other way around, we also compared the success per route, to see if the route walked accounted for the difference. However, this is not the case, as on both routes the success rate was 6.0.

Users' satisfaction. We measured the users' satisfaction by using the The Net Promoter Score (NPS). The NPS shows the balance between the number of users who are likely to promote a product and the ones who are likely to negatively affect a product (detractors). In our case, the product is a hypothetical app that implements one of the two auditory displays: the speech-based or the sonification-based one. To compute the NPS score, one deducts the percentage of promoters (users who gave scores of nine and 10) from the detractors (users who gave scores of six or below). The score runs from -100 to +100. The passive users (scores of seven or eight) are not considered at all. In our case, the NPS scores for both speech and sonification was zero - the number of promoters was equal to the one of detractors. The comments made to this question show that users were mostly dissatisfied with aspects related to the evaluation design. For instance, the high amount of information in the short time and the short pause between the instructions were the main reasons for criticism. Such issues would be rectified in a real app implementation, in which users can filter objects and change the speech rate and the pause between instructions. Two users made comments about the auditory icons being partly not intuitive (P5) or not being learned confidently during the evaluation (P4). Thus, having more time to get accustomed to the sounds and learn them confidently could increase users' satisfaction with such a system.

About the speech display, users liked that: "I can get a lot of valuable info" (P3); "If you can get the latest information from any place, the app can be very helpful" (P1); "good approach" (P4). Positive comments made about the sonification include: "It is very convenient to get a quick all-around view. The information can be easily visualized through the stereo image. The sounds are nice and short and the whole thing doesn't take long" (P1); "I can acquire a lot of important information in a very short time and thus orient myself directly in the building. By giving directions to objects even behind walls, I can see things that cannot be seen with my eyes. I can listen to the objects time and again, from any position, and thus check if I am going in the right direction" (P3).

Users' preference. When asked which of the two modalities they prefer, 3 users pleaded for the sonification and 2 for the speech instructions. This suggests that any of the two modalities can be used, depending on the user's preference, and more importantly, on the context of use. The comments made by participants to this question can be summarized in pros and cons for each auditory display - see Table 5.9.

Context of use for auditory icons. When asked in what context they would use sonification with auditory icons, the answers of the five participants were very diverse. We reproduce them in Table 5.10, together with a compact list of derived contexts of use.

Discussion. Both speech and sonification can be successfully used in interfaces for people with visual impairments. Which one is better depends on the context of use, the quality of the auditory icons used

Sonification	Speech
+ Faster + Learnable	+ Easier / clearer
- More difficult - Practice needed	- Slower

Table 5.9.: Advantages and disadvantages of sonification and speech, as mentioned in the comments of the 5 participants in the comparison evaluation.

	Participant quote	Context of use
P1	"Mainly to get an idea of things around me, but also specifically to find the exit/staircase. Rooms, on the other hand, are probably too many most of the time to learn anything meaningful about them."	Get an overview of the environment (P1, P5) Find things
P2	"Acoustic symbols are especially useful in environments that don't make much noise themselves. Acoustic symbols are better in quiet rooms or buildings."	Quiet environments
P3	"For information that is constantly changing and not necessarily in focus. So a kind of background noise that I can perceive when I need information, but otherwise can ignore without much problems to talk, for example. Then I find this kind of presentation easier than having everything read out in speech."	Dynamic (constantly changing) information As background information In parallel to conversations / speech interfaces
P4	"Structures with objects you expect; and familiar structures."	Familiar environments Typical environments, with predictable objects
P5	"Just get an impression of what is around me; but not for orientation / navigation."	Not for navigation / orientation

Table 5.10.: The context of use for auditory icons: the quotes of the 5 users on the left as well as the derived context of use, as compact list, on the right side of the table.

(and especially their intuitiveness) and users' preferences. We summarize some of the advantages and disadvantages of both speech and sonification in the following paragraphs, to support the decision of using one or the other.

The advantages of speech-based auditory displays include:

- Easy to understand, as long one is proficient in the language used;
- Requires minimal or no learning of the interface;
- Can convey almost *any* information, as long or as complex.

Advantages of sonification using auditory icons and parameter mapping include:

- More direct perception of the objects being conveyed;

- Shorter than speech at standard rate;
- Language-independent (but not necessarily culture-independent);
- Further characteristics of the objects, such as size, texture, temperature, etc., can be conveyed using other sound parameters. “For example, the pitch of the sound should change to indicate the direction of escalators and elevators and to distinguish the location of male and female toilets” (Bilal Salih et al. 2022).

Conclusion. Starting with a small expert round-table, we devised a method to convey acoustical information about objects in 3D space around the user using bone conducting headphones. The experts agreed on using panning for differentiating between left and right, and text-to-speech for differentiating between objects in front and objects behind the user. After that, in a pilot study with 5 users with visual impairments, we compared speech-based auditory displays with sonification-based ones using very short auditory icons and parameter mapping. Our evaluation showed that, for getting an overview of the environment, sonification is as good as speech in terms of *efficiency*, *effectiveness* and *user satisfaction*. The participants suggested several contexts in which they would like to use sonification displays based on auditory icons and parameter mapping. Thus, users would like to use sonification to get an overview of the environment, to search for things or as background information that can be used in parallel with speech. Some would only like to hear dynamic (changing) information, others would use sonification only in quiet or familiar environments. Finally, we present some of the advantages of both speech and sonification, to help with the decision making of choosing one or the other for an application.

6. Conclusion

6.1. Contributions

People with visual impairments are restricted in their mobility. While electronic travel aids have been in the focus of research for some time now, and new advances in artificial intelligence and computer vision make it possible to gather large amounts of information from the environment, the problem of the interface has been less addressed. In this thesis, we take a user-centered approach, involving users and experts in all stages of development - requirements gathering, prototyping and especially evaluation - to answer two main questions related to the auditory interface of electronic travel aids: (1) *What* information to convey to users? and (2) *How* to convey this information effectively?

Our contributions lie in the following three areas relating to people with visual impairments in the context of travel:

- Requirements for auditory displays (Chapter 3)
- Speech-based auditory displays (Chapter 4)
- Non-speech auditory displays or Sonification (Chapter 5)

In order to answer our research questions related to the interfaces of electronic travel aids for people with visual impairments, we conducted several studies, with both users and experts, in the 3 development phases: (1) requirements gathering; (2) prototyping and development; (3) evaluation. Table 6.1 summarizes the studies we performed. In total, we conducted **9 studies** with **60 users** with visual impairments and **32 experts**. We performed another pre-evaluation with 18 sighted people. Finally, 2 experts (one of whom was himself blind, thus also a user), were involved in the creation of the auditory icons, in a participatory development approach.

The most notable outcomes of this thesis, in short, are:

- **Requirements:** Several lists of indoor accessibility features, aligned with OpenStreetMap and the first step towards their standardization.
- **Requirements:** Set of non-functional requirements for accessible interfaces.
- **Interfaces:** Two grammars for conveying complex information needed by people with visual impairments in a structured way, one during outdoor navigation and orientation, and one during indoor orientation.
- **Interfaces:** A set of very short auditory icons for 40 everyday objects, both for indoor and outdoor use.

Section	**Development phase**	**Purpose**	**Type of study**	**With whom?**	**Number**
3.2.2	Requirements gathering	Assessment of requirements	Online questionnaire	Experts	22
4.1.1	Development	Creation of a German grammar for outdoors	Expert round-table	Experts	5
4.1.3	Evaluation	Evaluation of the speech auditory display for outdoors	On-site evaluation	Users	15
4.2.3	Evaluation	Evaluation of the speech auditory display for indoors	Online evaluation	Users	8
5.1	Proof of concept	Preliminary user study to asses if object sonification is feasible in the given context	On-site evaluation	Users	5
5.2.1	Requirements gathering	Determine conceptual mappings for auditory icons for a set of everyday objects	Online study	Users	11
5.2.2	Development	Creation of auditory icons	-	Experts	2
5.2.3	Evaluation	Pre-evaluation of the auditory icons	Online	Sighted team members	18
5.2.3	Evaluation	Evaluation of auditory icons	Online	Users	16
5.3	Design	How to design an auditory display on bone phones	Expert round-table, online	Experts	3
5.3	Evaluation	Comparison speech versus non-speech	On-site evaluation	Users	5
Total				**60 users** **32 experts** **18 sighted people**	

Table 6.1.: Summary of the studies performed in this thesis and the involvement of users and experts throughout all development stages during the design, development and evaluation of auditory displays for people with visual impairments in the context of travel.

- **Development:** Implementation of the interfaces in existing prototypes, as proof of concept and for evaluation purposes.
- **Evaluations:** Nine different studies were conducted with users and experts, both online, in the laboratory and on field, qualitative, quantitative (or both) and round tables. Their purpose was to gather requirements, design and asses the interfaces developed (grammars and set of auditory icons).

We present below in more detail our contributions in the field of auditory displays for people with visual impairments in the context of travel and answer the two main research questions.

Indoor accessibility features.
Related publication: Constantinescu et al. (2022a).
We assessed, by means of literature review, the *information needs* of people with disabilities in general and with visual impairments in particular, when travelling indoors. We compiled a large list of more than 820 entries for all disability types and a second list that addresses the needs of people with visual impairments only. To the best of our knowledge, this kind of information was never compiled before. Additionally, we compared the results found with the information that can be mapped in the free geographic database OpenStreetMap and suggested missing keys and tags. Finally, through the collaboration with SozialHelden, who are interested in further developing the lists, we set the cornerstone for standardization of accessibility features. Our efforts facilitate the introduction of the needed information in available maps that can be then used to create travel applications for people with disabilities.

Non-functional requirements for assistive output interfaces.
Through a literature review which included international and local (German) standards, international research papers and results of past projects conducted at our institution, we defined a set of non-functional requirements for the output interfaces of mobile assistive systems for people with visual impairments. Additionally, for each requirement, we defined a set of sample criteria. Through a subsequent expert questionnaire which was answered by 22 mobility trainers from Germany, we refined the set of requirements and assigned them an importance rating. All requirements except one were rated over 4 on a scale from 1 (not at all important) to 5 (very important), thus were considered to be important or very important. The only requirement that was considered less important was “affordability”, which was rated on average 3.7 (thus, above “neutral”). To be noted that the ratings might reflect the local situation and that in other countries the cost requirement might weigh more. Our requirements can be used both during the design phase of a mobile system, as well as a basis for heuristic evaluation.

Grammar for accessible outdoor travel.
Related publication: Constantinescu et al. (2019).
In a user-centered approach, starting with a round-table with five experts in mobility training, assistive technologies and linguistics, we defined a grammar for outdoor electronic travel aids. With this grammar, all the information needed by people with visual impairments during outdoor travel can be conveyed: navigation instructions that take into account the side of the road, the existence of sidewalks, landmarks and crossings with their signals and accessibility information. Additionally, a second syntax

(sentence structure) is used for conveying to users information about objects in the environment such as landmarks, points of interest or obstacles. We implemented the grammar into a prototype that features: (1) safe navigation guidance tailored to the needs of people with visual impairments (created by a project partner company) but with approximate distance estimates in cities provided by GPS; (2) precise additional information at crossings such as the existence and location of zebra crossings, traffic lights and their signal request button and the state of the traffic light (green versus red). This feature was implemented by project partners from KIT; (3) finally, a single, tailored speech interface based on our grammar that conveys all this information in a consistent way. Information from all sources is smoothly intertwined. We evaluated the prototype and with it, our grammar, with 15 people with visual impairments in an on-site study through the city. All participants preferred walking with the additional objects information and found this information useful. Suggestions on how to improve the interface were made.

Grammar for accessible indoor orientation.
Related publication: Constantinescu et al. (2022b).
Through literature review, we defined a German grammar for conveying information about objects in *indoor* environments. Subsequently, an online user study with 8 people with visual impairments was conducted in order to clarify a few aspects that remained ambiguous after the literature review. At the same time, the purpose of the user study was also to evaluate the created grammar. We show that the grammar is effective in helping users create a mental map of the environment. Moreover, users need and want to have a structured grammar with fixed word order in the sentences. We showed that a random word order frustrates the users and results in higher cognitive load and lower efficiency. Random word order is typical for natural language, as the most popular strategy used by people with visual impairments to get information when travelling to unknown buildings is to ask sighted passers by for directions. Based on the suggestions made by the participants in our user study, we propose improvements of our grammar as well as recommendations for systems implementing it. Finally, we implemented our grammar, in form of egocentric textual descriptions of the environment, into a prototypical App as a proof-of-concept.

Non-speech sonification for accessible travel.
Related publication: Constantinescu et al. (2020).
We created - in a user-centered way, involving experts and end users throughout development and evaluation, and carrying out 4 user studies - a *set of very short auditory icons for 40 everyday objects*. We first showed that *short* auditory icons combined with parameter mapping are suitable for conveying information about objects in the environment to people with blindness while travelling. Afterwards, we compiled a collection of conceptual mappings - sounds that blind people associate with 40 everyday objects. These can be used for creating new auditory icons. Subsequently, in a participatory development approach, we created a set of very short auditory icons for the 40 objects. The auditory icons have two versions: a 500 ms and a 1000 ms one. We then evaluated the shorter versions of the sounds with 16 blind users. The results showed that in most cases, users associate with the auditory icons either the correct object or some related concept. Many sounds were found to be pleasant and in few cases users made suggestions for improvements. Finally, we compared, in a user study with 5 people with visual impairments, speech with non-speech sonification in the context of

travel. Results show that for getting a rough overview of objects around, the chosen sonification is as efficient as text-to-speech.

We can summarize the answers to our main research questions as follows:

1. *What* information to provide?
    - We answer this question first of all in Section 3.1, in which we compile lists of indoor accessibility features. Secondly, we also answer this question in many of the various evaluations performed, in which users suggested what information they need in the given context.
    - The answer can be summarized as follows: people with visual impairments need a great amount of information. However, due to the heterogeneity of the group, each user needs different information in different contexts (indoor, outdoor, navigating to a target, exploring, searching). Thus, *provide as much information as possible*, but allow users to turn options on and off and configure.
2. *How* to convey the information?
    - We answered this question on the one hand in Section 3.2, but also in Chapters 4 and 5 in which we propose speech and non-speech interfaces. In the literature reviews and evaluations performed throughout these chapters, we also outline *how* to convey information to people with visual impairments during travel.
    - The answer to this question can be summarized as follows: besides the non-functional requirements defined in Section 3.2, we also found out that people with visual impairments need *structured* information. A grammar, for instance, is perfectly suitable for this purpose. Also, as mentioned above, the users should be able to configure the information provided according to their needs, preferences and context of use. For instance, they should be able to choose which object classes to convey and within what radius. This is in line with the non-functional requirement *Adjustability* - see Table 3.5.

Limitations. Our work addresses mainly the German population. This is first of all the case with the two grammars (Sections 4.1.1 and 4.2.2), which were specifically designed for the German language, but also the ratings of the non-functional requirements (Section 3.2). In other cases, such as the evaluations of the non-speech interface, even though we aimed to evaluate internationally, we received most answers from German-speaking countries or, at best, from the European area. Thus, our results might not apply to other cultures. Further evaluations will have to be performed to prove this. Another limitation is that the evaluations of the prototypes were mainly preliminary and often qualitative. To thoroughly evaluate our interfaces we need to implement them in mainstream apps and carry out long-term evaluations. This, however, was out of the scope of this thesis and remains as future work.

6.2. Future Work

As mentioned above, long-term evaluations are necessary for a thorough evaluation of the proposed interfaces. But not only the interfaces themselves should be evaluated this way. The comparison between speech and non-speech sonification should also be conducted in a long-term study within a real app. Also, we assume that most advantages come from a combination of speech with non-speech, but this should be proven in future work.

Another interesting direction is to find further adequate parameter mappings to map further object properties such as size, color, texture, and others. This will enable people with visual impairments to get even more information about their environment in a short time. Whether this is also intuitive enough, will have to be evaluated.

Regarding the indoor grammar, what the users in our studies wished was to get information about walls, the shapes of corridors and the layout of buildings or spaces in general. Users need this for creating a proper mental map, as they expressed it in the evaluations. Thus, one must find a way to convey this information acoustically, whether with speech or non-speech. Haptic interfaces should also be considered and combined with audio interfaces for best results.

Finally, it should be evaluated whether our interfaces are also appropriate for sighted people. Often, accessibility is a win for all, and it could be that sighted people will also profit from eyes-free interfaces. Looking at the smartphone's screen while walking on the street can be dangerous, thus, auditory displays might be the right solution in this case.

Our work can be used in navigation applications for people with visual impairments, for augmented and mixed reality, but also virtual reality (Kunz et al. 2018) and simulations, audio-haptic interactions such as on the TPad (Melfi et al. 2020) and games.

A. Appendix

A.1. Complete Outdoor Grammar

Table A.1 lists all the grammar rules we defined for various events on the route.

Ereignis	**Ansage Grammatik**	**Ansage Beispiel**
Start	"Route wird berechnet:" Distanz. "Start" Richtung Straße Straßenseite ["Richtung" Kreuzende Straße][1]	Route wird berechnet: 330 Meter. Start um 3 Uhr auf Schönfeldstraße, rechte Seite, Richtung Parkstraße.
Abbiegen, 25m	Distanz Straße Richtung Aktion	In 25 Metern auf Karl-Wilhelm-Platz scharf links abbiegen
Abbiegen, 10m	Distanz Richtung Aktion	In 10 Metern scharf links abbiegen
Überqueren, 25m	Distanz Straße Objekt* Richtung Aktion	In 25 Metern Karl-Wilhelm-Platz an der Ampel scharf links überqueren
Überqueren, 10m	Distanz Richtung Aktion	In 10 Metern rechts überqueren
Wechseln	Distanz Objekt Aktion	In 10/25 Metern Straßenseite wechseln
Folgen	Distanz Straße? Richtung? Aktion	Alle 25m: 25 Metern folgen Straße biegt ab: In 35m Karlstraße rechts folgen.
Folgen + Straßenseite[2]	Distanz Straße? Straßenseite Richtung? Aktion	75 Metern Schönfeldstraße auf linker Seite folgen (implizit: geradeaus)
Ziel, 25m	Distanz Straßenseite Objekt Aktion	In 25 Meter auf rechter/linker Seite Ziel erreicht
Beim Ziel	Objekt Aktion	Ziel erreicht
Tankstelle / Parkhaus	„Vorsicht!" Objekt Distanz Straßenseite	Vorsicht! Tankstelle/Parkhaus in 150 Meter auf der linken Seite
Kreuzungen beschreiben	Name. Typ-und-Anzahl-Straßen: [Richtung Straße Objekt*]X-mal	Karl-Wilhelm-Platz. 4-Wege-Kreisel: 2 Uhr Haid-und-Neu-Straße, mit Straßenbahnlinien; 6 Uhr Georg-Friedrich-Straße; 8 Uhr Karl-Wilhelm-Straße mit Straßenbahnlinien; 12 Uhr Parkstraße
Wo bin ich?	Standort: Straße Str-Nr., Stadt. Straßenseite. Richtung. Höhe. POIs Liste	Standort: Engesserstr. 4, Karlsruhe. Rechte Straßenseite. Ausrichtung: Nord-Ost. 115 Metern hoch. [Liste POIs]

Table A.1.: The rules for the outdoor grammar, in German, together with sample instructions.

---

[1] Optional

[2] Kann am Anfang der Route vorkommen, wenn das erste Ereignis "folgen" ist; oder nach einem Straßenseitenwechsel, auch in Verbindung mit dem Befehl "folgen"

A.2. Thematic Analysis of Conceptual Mappings

In the online questionnaire about the conceptual mappings of auditory icons, we asked the question: “What sounds best describe the following objects?” We then carried a thematic analysis on the answers. We list here the themes for all objects in the study.

Theme	Vol.	%	Type of object
Toilet or bathroom for men, women or people with disabilities			
Toilet flush	7	41,2%	
NO toilet flush - indiscreet, embarrassing	1	5,9%	
Washing hands - sink - rushing water - mountain spring	6	35,3%	
TTS: differentiated sex / people with disabilities / generic	2	11,8%	
Hum of neon lights	1	5,9%	
Elevator			
Humming of elevator engine	8	88,9%	
Typical "Ping" sound when elevator arrives	5	55,6%	
Elevator doors	3	33,3%	
TTS: "You are on the 1st floor"	1	11,1%	
The humming of bees that go into or out of the hive	1	11,1%	
Service point (reception, security office, service, information point, queue management ticket machine)			
People talking	4	26,7%	Service point
Bell, beep, bing, tri-tone	4	26,7%	Service point
Phone ringing	1	6,7%	Service point
Typing on a mechanical keyboard	1	6,7%	Service point
Ticket machine: ticking	1	6,7%	Ticket ma.
Room, room number and room type (office, conference room, laboratory, library, balcony)			
Different sounds	4	17%	
Combination of sounds: one for room + one for type of room		8%	
Door, wood	3	16,7%	
Muffled sound due to carpets	3	12,5%	
People talking	2	8,3%	
Key turning in the lock	1	4%	
No sound (2/3 TTS)	3	13%	
Eating place (canteen, cafeteria, coffee shop, kitchen, vending machine)			
Different sounds	2	10%	
Clattering of dishes and cuttlery	7	30%	
Glass	1	5%	
People, human voice rumble	3	13%	
A food processor/blender	1	4%	

A gong, as was once called for dinner	1	4%	
Kitchen: clattering of dishes	1	4%	Kitchen
Kitchen: dishwasher, oven or refrigerator	1	4%	Kitchen
Kitchen: frying and cooking	1	4%	Kitchen
Coffee Shop: Cups rattling on coasters	1	4%	Coffee shop
Vending machine: money falling into the machine	1	4%	Vending ma.
Vending machine: the rattle when the product is taken out or money	1	4%	Vending ma.
The hum of a vending machine's cooling	1	4%	Vending ma.
Floor number			
TTS announcement	5	42%	
Short sound (ding, honk)	3	25%	
Don't know	3	25%	
Steps on carpet or hard surface	1	8%	
Payment and cash withdrawal point (cashier‘s desk, ATM, cash machine)			
Different sounds	2	12%	
Beep (like in cash dispenser)	5	29%	
Coins falling	6	35%	
Cash dispenser noises	1	6%	Cash dispen.
A receipt being printed	1	6%	
The sound of the bill/banknotes reader	1	6%	
Money counter: conversation between employees and customers	1	6%	Money counter
Indoor pillar			
No sound	3	27%	
General synthetic warning signal for obstacles	1	9%	
Hitting pillar [cane, person, head]	5	45%	
A deep(low) whirring	1	9%	
Brick	1	9%	
Office electronics (equipment or machinery, computer, printer, photocopier)			
Different sounds	3	13%	
Typing	5	21%	
Ventilator hum	3	13%	
Printing	4	17%	Printer, Copy ma.
Beep	3	13%	
PC: mouse sounds	2	8%	
PC: Windows sounds, incl. startup	2	8%	
Copy machine: buzzing when a sheet is scanned	1	4%	Scanner, Copy ma.
Household appliance (refrigerator, television, washing machine)			
No sound, TTS	1	4%	

Different sounds	5	20%	
Hum, buzz	6	24%	
Refrigerator: humm, buzz	4	16%	Refrigerator
Refrigerator: fridge door	1	4%	Refrigerator
Refrigerator: rumble	1	4%	Refrigerator
Television: mumbled voices, music	4	16%	Television
Television: static noise	3	13%	Television
Washing machine: washing machine sounds	6	24%	Washing ma.
Washing maschine: noise of a mechanical toy	1	4%	Washing ma.
Vacuum cleaner	1	4%	

Electrical installation device (socket, light switch, intercom)

No sound	2	11%	
Different sounds	4	22%	
Buzz, humm, fizzle	3	17%	
Click when flipping switch	4	22%	[Light] Switch
Light switch: metallic click-crack	1	6%	[Light] Switch
Socket: the plug noise, light [scharen] and then a little louder "jolt" at the end	1	6%	Socket
Searching the holes of the socket with the metal pins of the cable	1	6%	Socket
Intercom: no sound [you hear it anyway, both inside and outside]	1	6%	Intercom
Intercom: a beep	1	6%	Intercom

Fire safety equipment (fire extinguisher, fire hose)

Different sounds	1	8%
No sound	1	8%
Siren, alarm	4	33%
Building fire sound	1	8%
Water	1	8%
"Dong" when you run into a fore extinguisher	1	8%
Don’t know	1	8%

Water source (sink or washbasin, shower, water dispenser, water tap)

Different sounds	3	21%
Running water, stream	6	43%
Tap sounds: running water, splash, water drops, person washing hands)	6	43%
Water drops	2	14%
Splashing from the tap	1	7%
Blink (flash)	1	7%

Heat source and hot items (radiator, heating panels, fireplace; heating burner; oven; hot water or anything that could be hot to burn)

Different sounds	1	8%
Fire crackling	6	46%
Loud bell/ high pitched sound	2	15%

Hiss of a teapot / high pitched sound	2	15%	
Heater: quiet [water] roaring in the heater	1	8%	Heater
Heater: unpleasant whistle and crackle	1	8%	Heater
Window (window, window shutters, rolling blinds, blinds)			
Different sounds	2	15%	
Wind	3	23%	
Glass (1/3 person rapping)	3	23%	
Window handle (open, close)	2	15%	
A window moved by the wind	1	8%	
Outside noise (birds, dog)	1	8%	
"Rip" (ratch) when pulling a curtain	1	8%	
Blinds: the ratchet of the cord and the blinds going up/down	2	15%	Blinds
Shutters: When closing, twice wood rattle	1	8%	Shutters
Handrail for indoor stairs, railings			
No sound	2	18%	
Different sounds	1	9%	
Metal (3/5 cane hits against railing)	5	45%	
A simultaneous low and high tone	1	9%	
Bricks	1	9%	
Furniture (sofa, couch, bed, wardrobe, closet, locker, hanging cabinet, cupboard, drawers, dresser, shelf, book case)			
No sound (1/3 TTS)	3	12%	
Different sounds	4	15%	
Wood (door, hangers)	6	23%	
Furniture door (creach, squeak, open, close)	5	19%	
Don't know	1	4%	
Table (table, desk, teacher‘s desk, tea table)			
No sound	1	6%	
Different sounds	4	22%	
Depending on what's on the table	1	6%	
[Knocking on] wood	5	28%	
Glass, teaspoon in a cup, plates	4	22%	
Something is placed on a large surface at stomach height	1	6%	
People sitting at distance to each other around something	1	6%	
Pencils rolling on a desk	1	6%	
Glass element (glass partition, glas surface, wall mirror)			
Different sounds	1	8%	
Don't know	1	8%	
Breaking glass, shards	4	33%	

Rapping, kicking on glass	3	33%	
Scrub	1	8%	
It sounds big when you walk towards it	1	8%	
Projecting wall			
No sound (TTS)	4	36%	
Head hitting against it	1	9%	
Tinkling, buzzing sound, like for electrical installation	2	18%	
People talking	1	9%	
Noise of power drill with percussion	1	9%	
Bicycle			
Ringing of bicycle bell	6	40%	
Rolling wheels	4	27%	
Chain (on gears)	3	20%	
The clacking of spokes attachments, like children have, to make funny noises when riding	1	7%	
Clang when you bump against it	1	7%	
Iron	1	7%	
Vehicle (motorcycle, car, truck, bus, train)			
Different sound for each type of vehicle	5	21,7%	
Engine, loud humming, roar	9	39,1%	All but train
Opening doors in public transport	3	13,0%	Public transp.
Tires	2	8,7%	Car, train
Motorcycle: accelerating engine	2	8,7%	
Car: engine, tires, horn	2	8,7%	
Truck: deep hum of engine; accelerating; beeping when it goes backwards	3	13,0%	
Bus: doors (swinging), engine	3	13,0%	
Train: breaks, tires, siren, doors	3	13,0%	
Train tracks			
Train, rail sound, distant	7	50,0%	
Metallic sound: gliding, walking over rails, iron, rumble, rattle	5	35,7%	
Barrier	1	7%	
Crosswalk (traffic light pole, zebra crossing)			
The sound made by accessible traffic lights to help the user find it	3	33,3%	
Metallic "ding" made by the white cane when hitting the traffic light pole	2	22,2%	
Cars	1	11%	
Steps	1	11%	
Stop sign (traffic sign)			
No sound	3	27,3%	

A high "bing" when the cane hits it; pole shaking	2	18,2%
Creaking breaks of an automobile as a warning	2	18%
Ticking, beeping	1	9%
Low bas rumble	1	9%

Fence

No sound	2	18%
Varying, depending on the material	1	9%
Metal fence: fast metal clang; shaking metal fence	3	27%
Wood fence: cane or wooden stick running along a picket fence (clack-clack-clack)	3	27%

Public transport station (train, metro, bus, taxi station)

TTS: name of station	2	13%
Different sounds	1	6,3%
Bus station: opening/closing doors, bus engine in the background	2	13%
Metro: opening/closing metro doors	1	6,3%
Train station: creak of trains stopping, with the acoustics of a large hall, if there is one	3	18,8%
Trams: trams stopping	1	6,3%
People talking	3	18,8%
Some public radio station sounding in the area	1	6,3%
You hear the canopy, if there is one	1	6,3%

Building

No sound	2	18,2%
Change of acoustics when the building is nearby	2	18,2%
Stone, wall, bricks	2	18,2%
Closing a creaching, metal door	2	18,2%
A two-part or better three-part sound to represent "3D"	1	9,1%

Pavement or sidewalk

Steps, people walking	4	40,0%
Cars with people moving along a straight line	1	10,0%
Rough rattling, rubbed with paper	2	20,00%
Pilot tone	1	10,00%
Wall	1	10,00%

Person

Voice, person talking	6	46,2%
Steps	3	23,1%
Breath	3	23,1%
Person sitting down with a sigh	1	7,7%

Dog, other animals

Sound of particular animal	5	31,3%
Bark	7	43,8%
Panting	1	6,3%
Clinking of metal objects hitting each other (dog leash, collar tag)	1	6,3%
Insects: wing beats	1	6,3%
Hard, short sound	1	6,3%

Door (simple, automatic, revolving, sliding, glass door, automatic bars, entrance or exit, emergency exit)

Different sounds [for different types of door]	3	16,67%	
No sound - it scares	1	5,56%	
TTS: door, entry	1	5,56%	
Emergency exit: no sound	1	5,56%	Emerg. exit
Sound for emergency with an alert	1	5,56%	Emerg. exit
Squeak, creach, wood	3	16,67%	Door
The bang when closing; snap when door falls into the lock	2	11,11%	Door
Key turning in the lock; door is unlocked	2	11,11%	Door
Keys	1	5,56%	Door
Sliding doors: [the sound] when opening and closing	1	5,56%	Sliding door
Automatic doors: the humming of engines; electric whirring	3	16,67%	Autom. door
Revolving doors: no	1	5,56%	Revolv. door

Stairs, steps, level change

Steps on stairs [Quick, high heels, concrete, going up]	7	46,67%
People [heavy breath, voices]	3	20,00%
Stone, bricks, concrete	3	20,00%
Different acoustics depending on width and condition	1	6,7%
Downwards: it opens/abyss; sound of people bellow	1	6,7%
Upwards: closed sound; sound of people above	1	6,7%

Obstacle (hanging or overhead; on the ground; short obstacle)

Warning beep	3	23,1%
No sound	3	23,1%
Differentiated (ground-high; what obstacle)	3	23,1%
Hitting [one's head] [against metal]	2	15,4%
Head-high obstacles: when a sound is blocked by the obstacle	1	5,9%
Description	1	7,7%

Seating (chair, sitting bench, armchair)

No sound	1	8%
Different sounds	2	15%
Don't know	1	8%
Chairs being dragged (wood or metal)	4	31%

"Bang"; cane hits wood or metal	3	23%
Someone lets himself fall noisily into the armchair	1	8%
Chairs and bench: people talking at chest height	1	8%

Ramp, slope

No sound (TTS)	2	16,7%
Person sliding, skis on a skislope, longer sound	3	25,0%
Stone, plastic	1	8,3%
Bricks	1	8,3%
The acoustics of a ramp (when you hear something in the distance that is higher up, but the sound comes more from the side. If there was a wall right next to you, the sound would come from directly above)	1	8,3%

Shop (clothing; electronics; food: supermarket, grocery shop, market place)

Different sounds, responding to their regular usage	1	6,7%
Beeping of scanners at the checkout counter (cashier)	4	26,7%
People, voices, conversation	4	26,7%
Shopping cart: clatter and roll	2	13,3%
A door bell, like from a store in the past, when you open the door	1	6,7%
Doors opening and closing	1	6,7%
Paper, glasses	1	6,7%

Dangerous point or area

No sound	4	33%
TTS: "caution", description	2	16,7%
Alarm, warning, SOS	3	25,0%
Harbor: water splashes, ship engines, ship horns	1	8,3%
Low multiple tone rumble	1	8,3%
Directional sound that indicates where the danger comes from	1	8,3%
Background music like in movies	1	8,3%

Trash can (garbage can, recycle bin)

Different sounds	2	12,5%
Metal (trash can, tin bucket)	2	12,5%
Something (paper, glass) being thrown in the trash	4	25,0%
Paper rustling, paper thrown in can	2	12,5%
(Opening and) Closing the trash lid	3	18,8%
Putting a plastic bag in the trash	3	18,8%
Slammed trashcans	1	6,3%

Tactile or haptic information (tactile paving; tactile signpost; haptic reference point or landmark; change of the surface or texture on the ground; carpet; floor condition)

No sound	2	18%
Different sounds	3	27%

Tactile paving: rattle, "crunchy" sound of cane moving over this irregular surface	4	36%	Tactile paving
Hand rubbing paper	1	9%	
White cane sounds dull	1	9%	Carpet
Plants (bush, flower pot)			
Different sounds	1	11%	
Wind	2	22%	
Rustling of (1xdry) leaves	4	44%	
White cane hitting a bush fence	1	11%	
Smoking area			
No sound (TTS)	2	18%	
(Stone) lighter sound	4	36%	
Puffing out the smoke	2	18%	

List of Publications

Angela Constantinescu, Vanessa Petrausch, Karin Müller, and Rainer Stiefelhagen. Towards a standardized grammar for navigation systems for persons with visual impairments. In *21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS, pages 539–541. ACM, 2019. doi: 10.1145/3308561.3354618.

Angela Constantinescu, Karin Müller, Monica Haurilet, Vanessa Petrausch, and Rainer Stiefelhagen. Bring the environment to life: A sonification module for people with visual impairments to improve situation awareness. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, ICMI, pages 50–59. Association for Computing Machinery, Inc, 2020. doi: 10.1145/3382507.3418874.

Angela Constantinescu, Karin Müller, Claudia Loitsch, Sebastian Zappe, and Rainer Stiefelhagen. Traveling to unknown buildings: accessibility features for indoor maps. In *International Conference on Computers Helping People with Special Needs*, ICCHP-AAATE, pages 221–228. Springer, 2022a.

Angela Constantinescu, Eva-Maria Neumann, Karin Müller, Gerhard Jaworek, and Rainer Stiefelhagen. Listening first: egocentric textual descriptions of indoor spaces for people with blindness. In *International Conference on Computers Helping People with Special Needs*, ICCHP-AAATE, pages 241–249. Springer, 2022b.

Christin Engel, Karin Müller, Angela Constantinescu, Claudia Loitsch, Vanessa Petrausch, Gerhard Weber, and Rainer Stiefelhagen. Travelling more independently: A requirements analysis for accessible journeys to unknown buildings for people with visual impairments. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS, pages 1–11. Association for Computing Machinery, Inc, 2020.

Other Own Publications

Most of the following publications were authored or co-authored by Angela Schön (nee Constantinescu), before her doctoral research. A few others were co-authored *during* the doctoral research, but are beyond the scope of this thesis.

Angela Constantinescu. *Interactiunea Om-Calculator - Suport didactic pentru coordonatorii educatiei incluzive*. Program de formare continua in domeniul educatiei incluzive. Inst. de Formare Continua, Chisinau, 2011.

Angela Constantinescu and Karin Müller. Sonification of ascii circuit diagrams. In Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer, editors, *Computers Helping People with Special Needs (ICCHP)*, Lecture Notes in Computer Science, pages 89–91, Berlin, Heidelberg, 2010. Springer.

Angela Constantinescu and Tanja Schultz. Redundancy versus complexity in auditory displays for object localization - a pilot study. In *Proceedings of the 17th International Conference on Auditory Display*, ICAD, Budapest, Hungary, 2011.

Manuel Martinez, Angela Constantinescu, Boris Schauerte, Daniel Koester, and Rainer Stiefelhagen. Cognitive evaluation of haptic and audio feedback in short range navigation tasks. In Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler, editors, *Computers Helping People with Special Needs*, pages 128–135, Cham, 2014. Springer International Publishing.

Manuel Martinez, Kailun Yang, Angela Constantinescu, and Rainer Stiefelhagen. Helping the Blind to Get through COVID-19: Social Distancing Assistant Using Real-Time Semantic Segmentation on RGB-D Video. *Sensors*, 20(18), 2020. doi: 10.3390/s20185202.

Karin Müller and Angela Constantinescu. Improving the accessibility of ascii graphics for the blind students: Producing my own graphics. In *Computers Helping People with Special Needs (ICCHP)*, Lecture Notes in Computer Science, page 439–442, Berlin, Heidelberg, 2010. Springer.

Sebastian Ritterbusch, Angela Constantinescu, and Volker Koch. Hapto-acoustic scene representation. In Klaus Miesenberger, Arthur Karshmer, Petr Penaz, and Wolfgang Zagler, editors, *Computers Helping People with Special Needs (ICCHP)*, volume 7383 of *Lecture Notes in Computer Science*, pages 644–650. Springer Verlag, 2012. doi: 10.1007/978-3-642-31534-3_94.

M. Saquib Sarfraz, Angela Constantinescu, Melanie Zuzej, and Rainer Stiefelhagen. A multimodal assistive system for helping visually impaired in social interactions. *Inform. Spektrum*, 40(6): 540–545, 2017. doi: 10.1007/s00287-017-1077-7.

Boris Schauerte, Manel Martinez, Angela Constantinescu, and Rainer Stiefelhagen. An assistive vision system for the blind that helps find lost things. In Klaus Miesenberger, Arthur Karshmer, Petr Penaz, and Wolfgang Zagler, editors, *Computers Helping People with Special Needs (ICCHP)*, pages 566–572, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

Jiaming Zhang, Kailun Yang, Angela Constantinescu, Kunyu Peng, Karin Mueller, and Rainer Stiefelhagen. Trans4trans: Efficient transformer for transparent object segmentation to help visually impaired people navigate in the real world. In *IEEE/CVF International Conference on Computer Vision Workshop*, ICCVW, 07 2021. doi: 10.1109/ICCVW54120.2021.00202.

Jiaming Zhang, Kailun Yang, Angela Constantinescu, Kunyu Peng, Karin Müller, and Rainer Stiefelhagen. Trans4trans: Efficient transformer for transparent object and semantic scene segmentation in real-world navigation assistance. *IEEE Transactions on Intelligent Transportation Systems*, 23 (10):19173–19186, 2022. doi: 10.1109/TITS.2022.3161141.

Bibliography

Richard Adams, Dianne Pawluk, Margaret Fields, and Ryan Clingman. Multiomodal application for the perception of spaces (MAPS). In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS)*, pages 393–394, New York, NY, USA, 2015. Association for Computing Machinery. doi: 10.1145/2700648.2811386.

AfterShokz. Trekz titanium. `https://aftershokz.com/products/trekz-titanium`, 2020. Accessed: February 03, 2020.

Dragan Ahmetovic, Federico Avanzini, Cristian Bernareggi, Gabriele Galimberti, Luca Ludovico, Sergio Mascetti, and Giorgio Presti. Sonification of rotation instructions to support navigation of people with visual impairment. pages 1–10, 2019. doi: 10.1109/PERCOM.2019.8767407.

Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Marco Ciardullo, Gabriele Galimberti, Luca A. Ludovico, Sergio Mascetti, and Giorgio Presti. Sonification of navigation instructions for people with visual impairment. *International Journal of Human-Computer Studies*, 177, 2023. doi: 10.1016/J.IJHCS.2023.103057.

Aipoly Inc. Aipoly vision: Sight for blind & visually impaired. `https://apps.apple.com/us/app/aipoly-vision-sight-for-blind-visually-impaired/id1069166437`, 2020. Accessed: August 8, 2020.

Shurug Al-Khalifa and Muna Al-Razgan. Ebsar: Indoor guidance for the visually impaired. *Computers & Electrical Engineering*, 54:26–39, 2016. doi: 10.1016/j.compeleceng.2016.07.015.

Abdulrhman Alkhanifer and Stephanie Ludi. Towards a situation awareness design to improve visually impaired orientation in unfamiliar buildings: Requirements elicitation study. In *Requirements Engineering Conference (RE)*. Institute of Electrical and Electronics Engineers (IEEE), 2014. doi: 10.1109/RE.2014.6912244.

Abdulrhman A Alkhanifer. *The role of situation awareness metrics in the assessment of indoor orientation assistive technologies that aid blind individuals in unfamiliar indoor environments*. PhD thesis, 2015.

Gary L Allen. Principles and practices for communicating route knowledge. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 14(4):333–359, 2000.

Apple Inc. Maps. A better way to get there. `http://www.apple.com/ios/maps`, 2018. Accessed: September 21, 2018.

Malika Auvray, Sylvain Hanneton, and J Kevin O'Regan. Learning to perceive with a visuo - auditory substitution system: localisation and object recognition with 'The Voice'. *Perception*, 36(3):416–430, 2007. doi: 10.1068/p5631.

Nida Aziz, Tony Stockman, and Rebecca Stewart. An investigation into customisable automatically generated auditory route overviews for pre-navigation. In *Proceedings of the International Conference on Auditory Display (ICAD)*, 2019. doi: 10.21785/icad2019.029.

Jeffrey Baake. Google Play. Nav by ViaOpta - Bewertungen und Rezensionen. `https://play.google.com/store/apps/details?id=com.novartis.blind&hl=de&pli=1`, oct 2020.

Jan Balata, Jakub Berka, and Zdenek Mikovec. Indoor-outdoor intermodal sidewalk-based navigation instructions for pedestrians with visual impairments. In Klaus Miesenberger and Georgios Kouroupetroglou, editors, *Computers Helping People with Special Needs*, pages 292–301. Springer International Publishing, 2018.

Michael Banf and Volker Blanz. A modular computer vision sonification model for the visually impaired. In *The 18th International Conference on Auditory Display (ICAD)*, 2012.

Aaron Bangor, Phil T Kortum, and James T Miller. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24:574–594, 2008. doi: 10.1080/10447310802205776.

Nikola Banovic, Rachel L. Franz, Khai N. Truong, Jennifer Mankoff, and Anind K. Dey. Uncovering information needs for independent spatial learning for users who are visually impaired. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, New York, NY, USA, 2013. Association for Computing Machinery. doi: 10.1145/2513383.2513445.

Janet M. Barlow, Billie Louise Bentzen, and Tamara Bond. Blind pedestrians and the changing technology and geometry of signalized intersections: Safety, orientation, and independence. *Journal of visual impairment & blindness*, 99, 2005.

Fereshteh S. Bashiri, Eric LaRose, Jonathan C. Badger, Roshan M. D'Souza, Zeyun Yu, and Peggy Peissig. Object detection to assist visually impaired people: A deep neural network adventure. In George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Matt Turek, Srikumar Ramalingam, Kai Xu, Stephen Lin, Bilal Alsallakh, Jing Yang, Eduardo Cuervo, and Jonathan Ventura, editors, *Advances in Visual Computing*, pages 500–510, Cham, 2018. Springer International Publishing.

Durand Begault, Martine Godfroy-Cooper, Aniko Sandor, and Kritina Holden. Auditory alarm design for NASA CEV applications. In *Proceedings of the International Conference on Auditory Display*, ICAD, 2012.

Berufsförderungswerk Würzburg. BFW SmartInfo. `smartinfo.bfw-wuerzburg.de`. Accessed: April 19, 2023.

Alexy Bhowmick and Shyamanta M Hazarika. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *Journal on Multimodal User Interfaces*, 11(2):149–172, 2017.

Brandon Biggs, James M. Coughlan, and Peter Coppin. Design and evaluation of an audio game-inspired auditory map interface. In *Proceedings of the 25th International Conference on Auditory Display*, ICAD, 2019.

Hisham E. Bilal Salih, Kazunori Takeda, Hideyuki Kobayashi, Toshibumi Kakizawa, Masayuki Kawamoto, and Keiichi Zempo. Use of auditory cues and other strategies as sources of spatial information for people with visual impairment when navigating unfamiliar environments. *International Journal of Environmental Research and Public Health*, 19(6), 2022. doi: 10.3390/ijerph19063151.

Karin Böhm. Gesundheit. Auszug aus dem Datenreport 2021. Technical report, Statistisches Bundesamt (Destatis), 2021. `https://www.destatis.de/DE/Service/Statistik-Campus/Datenreport/Downloads/datenreport-2021-kap-9.pdf?__blob=publicationFile`.

Stacy M. Branham and Shaun K. Kane. The invisible work of accessibility: How blind employees manage accessibility in mixed-ability workplaces. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS)*, pages 163–171, New York, NY, USA, 2015. Association for Computing Machinery. doi: 10.1145/2700648.2809864.

Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006. doi: 10.1191/1478088706qp063oa.

Annette Bredmose, Sidse Grangaard, Victoria Linn Lygum, and Anders Rhiger Hansen. Mapping the importance of specific physical elements in urban space for blind and visually impaired people. *Journal of Urban Design*, 28(2):139–154, 2023. doi: 10.1080/13574809.2022.2097862.

Anke M. Brock, Philippe Truillet, Bernard Oriola, Delphine Picard, and Christophe Jouffrais. Interactivity improves usability of geographic maps for visually impaired people. *Human–Computer Interaction*, 30(2):156–194, 2015. doi: 10.1080/07370024.2014.924412.

John Brooke. SUS: a retrospective. *Journal of Usability Studies*, 8:29–40, 2013.

Lars Büchel. Peas at 5:30. `https://www.imdb.com/title/tt0366416/`, 2004.

Bundesministerium der Justiz. Anlage zu Absatz 2 der Versorgungsmedizin-Verordnung vom 10. Dezember 2008. Anlage “Versorgungsmedizinische Grundsätze”. Technical Report G5702, Bonn, 2008.

Bundesministerium für Verkehr, Bau und Stadtentwicklung und Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit. Verordnung zur Neufassung der Straßenverkehrs-Ordnung (StVO). *Bundesgesetzblatt online*, I(12):367–427, 2013.

João Paulo Cabral and Gerard Bastiaan Remijn. Auditory icons: Design and physical characteristics. *Applied Ergonomics*, 78:224–239, 2019. doi: https://doi.org/10.1016/j.apergo.2019.02.008.

Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

Fabio Castelli. Gespeaker, 2020. URL `http://www.muflone.com/gespeaker/english`.

Narciso Cerpa and June M. Verner. Why did your project fail? *Communications of the ACM*, 52(12): 130–134, 2009. doi: 10.1145/1610252.1610286.

Konstantinos Charitakis and Konstantinos Papadopoulos. ATMAPS. User Requirements Specifications Report. Technical Report D2.1, Project ATMAPS: Specification of symbols used on audio-tactile maps for individuals with blindness, 2017. `https://www.atmaps.eu/deliverables/ATMAPS-D_2_1-User_requirements_and_specifications_report.pdf`.

Bettye Rose Connell, Mike Jones, Ron Mace, Jim Mueller, Abir Mullick, Elaine Ostroff, Jon Sanford, Ed Steinfeld, Molly Story, and Gregg Vanderheiden. The principles of universal design. `https://projects.ncsu.edu/design/cud/about_ud/udprinciplestext.htm`, 1997. Accessed: September 16, 2021.

John J Correia, Jeffrey M Stefan, and Jasmin Jijina. Method for providing vehicle navigation instructions, 2002. US Patent 6,424,912.

Dimitrios Dakopoulos and Nikolaos G Bourbakis. Wearable obstacle avoidance electronic travel aids for blind : A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(1):25–35, 2010.

Dictionary by Merriam-Webster. America's most-trusted online dictionary. intuitive. Definition of intuitive. `https://www.merriam-webster.com/dictionary/intuitive`, 2020.

Paul DiMaggio. Culture and cognition. *Annual review of sociology*, 23(1):263–287, 1997.

DIN-Fachbericht 124. DIN-Fachbericht 124 Gestaltung barrierefreier Produkte. Standard DIN-Fachbericht 124, DIN Deutsches Institut für Normung e.V., Berlin, Wien, Zürich, 2002.

Tilman Dingler, Jeffrey Lindsay, and Bruce N. Walker. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the 14th International Conference on Auditory Display (ICAD)*, 2008.

Alan Dix, Janet Finlay, Gregory D. Abowd, and Russell Beale. *Human-Computer Interaction*. Pearson Education Limited, 3rd edition, 2004.

Florian Dramas, Bernard Oriola, Brian G. Katz, Simon J. Thorpe, and Christophe Jouffrais. Designing an assistive device for the blind based on object localization and augmented auditory reality. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 263–264, New York, NY, USA, 2008. Association for Computing Machinery. doi: 10.1145/1414471.1414529.

Alistair D.N. Edwards. *Speech Synthesis: Technology for Disabled People*. Paul Chapman Educational Publishing, 1991. ISBN 978-1853960666.

M. Angeles Espinosa, Simon Ungar, Esperanza Ochaita, Mark Blades, and Christopher Spencer. Comparing methods for introducing blind and visually impaired people to unfamiliar urban environments. *Journal of Environmental Psychology*, 18(3):277–287, 1998.

European Telecommunications Standards Institute. Human Factors (HF); Guidelines for the design of mobile ICT devices and their related applications for people with cognitive disabilities. Standard ETSI EG 203 350 V1.1.1, European Telecommunications Standards Institute (ETSI), Sophia Antipolis Cedex, 2016.

Monica Fabiani, Victoria A. Kazmerski, Yael M. Cycowicz, and David Friedman. Naming norms for brief environmental sounds: Effects of age and dementia. *Psychophysiology*, 33(4):462–475, 1996. doi: https://doi.org/10.1111/j.1469-8986.1996.tb01072.x.

Navid Fallah, Ilias Apostolopoulos, Kostas Bekris, and Eelke Folmer. The user as a sensor: Navigating users with visual impairments in indoor spaces using tactile landmarks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI, pages 425–432, New York, NY, USA, 2012. Association for Computing Machinery.

Navid Fallah, Ilias Apostolopoulos, Kostas Bekris, and Eelke Folmer. Indoor Human Navigation Systems: A Survey. *Interacting with Computers*, 25(1):21–33, 2013. doi: 10.1093/iwc/iws010.

Mexhid Ferati, Steve Mannheimer, and Davide Bolchini. Usability evaluation of acoustic interfaces for the blind. In *Proceedings of the 29th ACM International Conference on Design of Communication*, SIGDOC '11, pages 9–16, New York, NY, USA, 2011. Association for Computing Machinery. doi: 10.1145/2038476.2038479.

Alexander Fiannaca, Ilias Apostolopoulous, and Eelke Folmer. Headlock: A wearable navigation aid that helps blind cane users traverse large open spaces. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS)*, pages 19–26, New York, NY, USA, 2014. Association for Computing Machinery. doi: 10.1145/2661334.2661453.

Free Software Foundation, Inc. eSpeak text to speech, 2020. URL `http://espeak.sourceforge.net`.

Ronja Frimalm, Johan Fagerlönn, Stefan Lindberg, and Anna Sirkka. How many auditory icons, in a control room environment, can you learn? In *Proceedings of the International Conference on Auditory Display*, ICAD, 2014.

Lukas Fritsch. Erstellung von Auditory Icons für Indoor Objekte für Menschen mit Blindheit. Bachelor's thesis, Karlsruhe Institute of Technology, Department of Informatics, Karlsruhe, 2020.

Funka Nu AB. Guidelines for the development of accessible mobile interfaces. Technical Report WCAG 2.1, Funka Nu., Stockholm, n.d.

Thomas Gallagher, Elyse Wise, Hoe Chee Yam, Binghao Li, Euan Ramsey-Stewart, Andrew G. Dempster, and Chris Rizos. Indoor navigation for people who are blind or vision impaired: where are we and where are we going? *Journal of Location Based Services*, 8(1):54–73, 2014. doi: 10.1080/17489725.2014.895062.

Florence Gaunet and Xavier Briffault. Exploring the functional specifications of a localized wayfinding verbal aid for blind pedestrians: Simple and structured urban areas. *Human–Computer Interaction*, 20(3):267–314, 2005. doi: 10.1207/s15327051hci2003_2.

William Gaver. The sonicfinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4:67–94, 1989. doi: 10.1207/s15327051hci0401_3.

William W. Gaver. Auditory icons: Using sound in computer interfaces. *Human–Computer Interaction*, 2(2):167–177, 1986. doi: 10.1207/s15327051hci0202_3.

Geluid in Zicht Foundation. Acoustic scale model. `http://www.geluidinzicht.nl/acoustic-scale-model/`, 2014. Accessed: October 10, 2023.

Michele Geronazzo, Alberto Bedin, Luca Brayda, Claudio Campus, and Federico Avanzini. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, 85:4–15, 2016. doi: 10.1016/j.ijhcs.2015.08.004. Data Sonification and Sound Design in Interactive Systems.

GMA Games. ShadesOfDoom. `https://www.audiogames.net/db.php?id=shadesofdoom`, 2020.

Mira Goldschmidt. *Mobility of Visually Impaired People: Fundamentals and ICT Assistive Technologies*, chapter Orientation and Mobility Training to People with Visual Impairments, pages 237–261. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-54446-5_8.

Reginald G. Golledge, James R. Marston, and C. Michael Costanzo. Attitudes of visually impaired persons toward the use of public transportation. *Journal of Visual Impairment & Blindness*, pages 446–459, 1997.

Jimena Gomez, M. Patrick Langdon, and John P. Clarkson. Navigating the workplace environment as a visually impaired person. In *Universal Access in Human-Computer Interaction*, UAHCI, 2016.

Google. Google maps. `http://www.google.com/maps`, 2018. Accessed: September 21, 2018.

Shreya Goyal, Satya Bhavsar, Shreya Patel, Chiranjoy Chattopadhyay, and Gaurav Bhatnagar. Sugaman: describing floor plans for visually impaired by annotation learning and proximity-based grammar. *IET Image Processing*, 13(13):2623–2635, 2019.

Rebecca Grier. How high is high? A metanalysis of NASA TLX global workload scores. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, 2015. doi: 10.1177/1541931215591373.

Cornelia Grohmann, Elisa Rudolph, and Carolina Kolodziej. *Kriterienkatalog zur Bewertung der baulichen Inklusion an Hochschulen*. Sozial- und Gesundheitsbauten, Professur für Gebäudelehre und Entwerfen der Universität Dresden, 2018. Accessed: June 18, 2021.

João Guerreiro and Daniel Gonçalves. Faster text-to-speeches: Enhancing blind people's information scanning with faster concurrent speech. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers Accessibility (ASSETS)*, 10 2015.

Maya Gupta, Ali Abdolrahmani, Emory Edwards, Mayra Cortez, Andrew Tumang, Yasmin Majali, Marc Lazaga, Samhitha Tarra, Prasad Patil, Ravi Kuber, and Stacy M. Branham. Towards more universal wayfinding technologies: Navigation preferences across disabilities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3313831.3376581.

Sandra G. Hart. Nasa-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. doi: 10.1177/154193120605000909.

Shirin E. Hassan. Are normally sighted, visually impaired, and blind pedestrians accurate and reliable at making street crossing decisions? *Investigative Ophthalmology & Visual Science*, 53(6):2593–2600, 2012.

Thomas Hermann. *Sonification for Exploratory Data Analysis*. PhD thesis, Universität Bielefeld, 2002.

Thomas Hermann, Andy Hunt, and John G. Neuhoff, editors. *The Sonification Handbook*. Logos Publishing House, Berlin, Germany, 2011. `http://sonification.de/handbook`.

Marion Hersh. Mental maps and the use of sensory information by blind and partially sighted people. *ACM Transactions on Accessible Computing*, 13(2), 2020. doi: 10.1145/3375279.

Marion A. Hersh. *Mobility of Visually Impaired People: Fundamentals and ICT Assistive Technologies*, chapter Mobility Technologies for Blind, Partially Sighted and Deafblind People: Design Issues, pages 377–409. Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-54446-5_13.

Rebecca J. Hirst, Lucy Cragg, and Harriet A. Allen. Vision dominates audition in adults but not children: A meta-analysis of the colavita effect. *Neuroscience & Biobehavioral Reviews*, 94: 286–301, 2018. doi: https://doi.org/10.1016/j.neubiorev.2018.07.012.

Monika Holfeld. *Barrierefreie Lebensräume: Bauen and Wohnen ohne Hindernisse. 2. überarbeitete Auflage*. Beuth. DIN e.V., 2011.

Haosheng Huang and Georg Gartner. A survey of mobile indoor navigation systems. In Ortag F. Gartner Georg, editor, *Cartography in Central and Eastern Europe. Lecture Notes in Geoinformation and Cartography*, pages 305–319. Springer, Berlin, Heidelberg, 2009. doi: 10.1007/978-3-642-03294-3_20.

Andreas Hub, Joachim Diepstraten, and Thomas Ertl. Design and development of an indoor navigation and object identification system for the blind. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 147–152, New York, USA, 2003. Association for Computing Machinery. doi: 10.1145/1028630.1028657.

Impression Games. CaesarIII. `https://de.wikipedia.org/wiki/Caesar_(Spieleserie)#Caesar_III`, 2020. Accessed: August 8, 2020.

Interaction Design Foundation. I like, i wish, what if. `https://public-media.interaction-design.org/pdf/I-Like-I-Wish-What-If.pdf`, 2020. Accessed: August 8, 2020.

ISO 9241-171:2008. Ergonomie der Mensch-System-Interaktion - Teil 171: Leitlinien für die Zugänglichkeit von Software; Deutsche Fassung. Standard, International Organization for Standardization, Berlin, 2008.

ISO/IEC Joint Technical Advisory Group. Guide for addressing accessibility in standards. Standard ISO/IEC GUIDE 71:2014(E), International Organization for Standardization and International Electrotechnical Commission, Geneva, 2014. Second edition.

ISO/TS 9241-126:2019. Ergonomics of human-system interaction - part 126: Guidance on the presentation of auditory information. Standard ISO/TS 9241-126:2019, International Organization for Standardization, Geneva, CH, dec 2019.

Rosen Ivanov. An approach for developing indoor navigation systems for visually impaired people using Building Information Modeling. *Journal of Ambient Intelligence and Smart Environments (JAISE)*, 9(4):449–467, 2017. doi: 10.3233/AIS-170441.

Christophe Jacquet, Yacine Bellik, and Yolaine Bourda. A context-aware locomotion assistance device for the blind. In Sally Fincher, Panos Markopoulos, David Moore, and Roy Ruddle, editors, *People and Computers XVIII – Design for Life*, volume 3118, pages 438–445, London, 2004. Springer London. doi: 10.1007/1-84628-062-1_20.

Dhruv Jain. Pilot evaluation of a path-guided indoor navigation system for visually impaired in a public museum. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS)*, pages 273–274, New York, NY, USA, 2014a. Association for Computing Machinery. doi: 10.1145/2661334.2661405.

Dhruv Jain. Path-guided indoor navigation for the visually impaired using minimal building retrofitting. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS)*, pages 225–232, New York, NY, USA, 2014b. ACM. doi: 10.1145/2661334.2661359.

Petra Jaunich. *m4guide - mobile multi-modal mobility guide AP 820. Auswertung der Befragung zur Ausgangssituation und den Anforderungen der Nutzer*. BLIC, Januar 2014. `http://docplayer.org/7733335-M4guide-mobile-multi-modal-mobility-guide-ap-820-auswertung-der-befragung-zur-ausgangssituation-und-den-anforderungen-der-nutzer.html`.

Watthanasak Jeamwatthanachai, Michael Wald, and Gary Wills. Indoor navigation by blind people: behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment*, 5:140–153, 2019.

John M Evans LLC. Standards for visual acuity. Technical report, National Institute for Standards and Technology, Newtown, CT, 2006.

Lynette A. Jones. *Encyclopedia of Neuroscience*, chapter Haptics, pages 1808–1811. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. doi: 10.1007/978-3-540-29678-2_2140.

Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Hironobu Takagi, and Chieko Asakawa. Supporting orientation of people with visual impairment: Analysis of large scale usage data. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 151–159, 2016.

Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Valeria Alampi, Hironobu Takagi, and Chieko Asakawa. Insights on assistive orientation and mobility of people with visual impairment based on large-scale longitudinal data. *ACM Transactions on Accessible Computing*, 11, 2017. doi: 10.1145/3178853.

Hernisa Kacorri, Eshed Ohn-Bar, Kris M. Kitani, and Chieko Asakawa. Environmental factors in indoor navigation based on real-world trajectories of blind users. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3173574.3173630.

Fotis P. Kalaganis, Panagiotis Migkotzidis, Kostas Georgiadis, Elisavet Chatzilari, Spiros Nikolopoulos, and Ioannis Kompatsiaris. Lending an artificial eye: Beyond evaluation of CV-based assistive systems for visually impaired people. In Margherita Antona and Constantine Stephanidis, editors, *Universal Access in Human-Computer Interaction. Access to Media, Learning and Assistive Environments*, pages 385–399, Cham, 2021. Springer International Publishing.

Rie Kamikubo, Naoya Kato, Keita Higuchi, Ryo Yonetani, and Yoichi Sato. Support strategies for remote guides in assisting people with visual impairments for effective indoor navigation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI 2020, pages 1–12, New York, NY, USA, 2020. ACM. doi: 10.1145/3313831.3376823.

Slim Kammoun, Florian Dramas, Bernard Oriolaand, and Christophe Jouffrais. Route selection algorithm for blind pedestrian. In *International Conference on Control, Automation and Systems*, pages 2223–2228, 2010. doi: 10.1109/ICCAS.2010.5669846.

Roya Norouzi Kandalan and Kamesh Namuduri. Techniques for constructing indoor navigation systems for the visually impaired: A review. *IEEE Transactions on Human-Machine Systems*, 50 (18):492–506, 2020. doi: 10.1186/s13673-020-00222-0.

Brian Katz, Slim Kammoun, Gaëtan Parseihian, Olivier Gutierrez, Adrien Brilhault, Malika Auvray, Philippe Truillet, Michel Denis, Simon Thorpe, and Christophe Jouffrais. Navig: Augmented reality guidance system for the visually impaired: Combining object localization, gnss, and spatial audio. *Virtual Reality*, 16:17, 01 2012.

Seita Kayukawa, Keita Higuchi, João Guerreiro, Shigeo Morishima, Yoichi Sato, Kris Kitani, and Chieko Asakawa. BBeep: A sonic collision avoidance system for blind travellers and nearby pedestrians. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2019. Association for Computing Machinery. doi: 10.1145/3290605.3300282.

Jee-Eun Kim, Masahiro Bessho, Shinsuke Kobayashi, Noboru Koshizuka, and Ken Sakamura. Navigating visually impaired travelers in a large train station using smartphone and bluetooth low energy.

In *Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC)*, pages 604–611, New York, NY, USA, 2016. Association for Computing Machinery. doi: 10.1145/2851613.2851716.

Andrew Kirkpatrick, Joshue O Connor, Alastair Campbell, and Michael (Ed.) Cooper. Web content accessibility guidelines (wcag) 2.1. w3c editor‘s draft. Technical Report WCAG 2.1, W3C, 2018. `https://www.w3.org/TR/WCAG21/`.

Kish, Daniel. Flashsonar: Understanding and applying sonar imaging to mobility. `https://nfb.org/sites/default/files/images/nfb/publications/fr/fr30/1/fr300107.htm`, 2011. Accessed: June 4, 2023.

KIT and IOSB. Das System SmartCampus und das Projekt SmartCampus barrierefrei. `https://cm.tm.kit.edu/smartcampus.php`, 2015. Accessed: June 18, 2021.

Barbara Ann Kitchenham and Stuart Charters. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, 07 2007.

Dennis H. Klatt. Review of text-to-speech conversion for english. *The Journal of the Acoustical Society of America*, 82(3):737–93, 1987.

Eunjeong Ko, Jin Sun Ju, and Eun Yi Kim. Situation-based indoor wayfinding system for the visually impaired. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 35–42, New York, NY, USA, 2011. Association for Computing Machinery. doi: 10.1145/2049536.2049545.

Daniel Koester. *An orientation mobility aid for people with visual impairments*. PhD thesis, Karlsruhe Institute of Technology, Department of Informatics, 2020.

komoot GmbH. Your route to adventure. `https://www.komoot.com/`, 2022. Accessed: July 29, 2022.

Philip Kortum. *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*. Morgan Kaufmann, 06 2008.

Athanasios Koutsoklenis and Konstantinos Papadopoulos. Haptic cues used for outdoor wayfinding by individuals with visual impairments. *Journal of visual impairment & blindness*, 18:43–53, 2014. doi: 10.1177/0145482X1410800105.

Arni Kristjansson, Alin Moldoveanu, Omar I. Johannesson, Oana Balan, Simone Spagnol, Vigdis Vala Valgeirsdottir, and Runar Unnthorsson. Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative Neurology and Neuroscience*, 34(5), 2016. doi: 10.3233/RNN-160647.

Vladimir A. Kulyukin, John Nicholson, David A. Ross, James R. Marston, and Florence Gaunet. The blind leading the blind: Toward collaborative online route information management by individuals with visual impairments. In *Social Information Processing, Papers from the 2008 AAAI Spring Symposium, Technical Report SS-08-06, Stanford, California, USA*, pages 54–59. AAAI, 2008.

Dumisani Kunene, Hima Vadapalli, and Jaco Cronje. Indoor sign recognition for the blind. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists (SAICSIT)*, New York, NY, USA, 2016. Association for Computing Machinery. doi: 10.1145/2987491.2987530.

Andreas Kunz, Klaus Miesenberger, Limin Zeng, and Gerhard Weber. Virtual navigation environment for blind and low vision people. In Klaus Miesenberger and Georgios Kouroupetroglou, editors, *Computers Helping People with Special Needs*, pages 114–122, Cham, 2018. Springer International Publishing.

Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. Tools and technologies for blind and visually impaired navigation support: A review. *IETE Technical Review*, 39(1):1–16, 2020. doi: 10.1080/02564602.2020.1819893.

Stephen Lane, Paidi O'Raghallaigh, and David Sammon. Requirements gathering: the journey. *Journal of Decision Systems*, 25(sup1):302–312, 2016. doi: 10.1080/12460125.2016.1187390.

Chen-Fu Liao. Using a smartphone app to assist the visually impaired at signalized intersections. Technical report, Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota, 2012.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 2014.

Anuradha Madugalla, Kim Marriott, Simone Marinai, Samuele Capobianco, and Cagatay Goncu. Creating accessible online floor plans for visually impaired readers. *ACM Transactions on Accessible Computing*, 13(4), 2020. doi: 10.1145/3410446.

Liz Mariner. *Cleared for Takeoff: English for Pilots*. A-E-Link Publications, Limited, 2007. ISBN 9780979506802.

James R Marston, Reginald G Golledge, and C Michael Costanzo. Investigating travel behavior of nondriving blind and vision impaired people: The role of public transit. *The Professional Geographer*, 49(2):235–245, 1997.

Sergio Mascetti, Lorenzo Picinali, Andrea Gerino, Dragan Ahmetovic, and Cristian Bernareggi. Sonification of guidance data during road crossing for people with visual impairments or blindness. *International Journal of Human-Computer Studies*, 85:16–26, 2016. doi: 10.1016/j.ijhcs.2015.08.003. Data Sonification and Sound Design in Interactive Systems.

Keenan R. May, Brianna J. Tomlinson, Xiaomeng Ma, Phillip Roberts, and Bruce N. Walker. Spotlights and soundscapes: On the design of mixed reality auditory environments for persons with visual impairment. *ACM Transactions on Accessible Computing*, 13(2):1–47, 2020. doi: 10.1145/3378576.

Lesley McIntyre. *The way-finding journey within a large public building: a user centred study of the holistic way-finding experience across a range of visual ability*. PhD thesis, Univ. Dundee, 2011.

M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan. A compressive sensing approach to describe indoor scenes for blind people. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(7):1246–1257, 2015.

Mohamed L. Mekhalfi, Farid Melgani, Yakoub Bazi, and Naif Alajlan. Fast indoor scene description for blind people with multiresolution random projections. *Journal of Visual Communication and Image Representation*, 44:95–105, 2017. doi: 10.1016/j.jvcir.2017.01.025.

Giuseppe Melfi, Karin Müller, Thorsten Schwarz, Gerhard Jaworek, and Rainer Stiefelhagen. Understanding what you feel: A mobile audio-tactile system for graphics used at schools with students with visual impairment. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, New York, NY, USA, 2020. Association for Computing Machinery. doi: 10.1145/3313831.3376508.

Mei Miao, Martin Spindler, and Gerhard Weber. Requirements of indoor navigation system from blind users. In Andreas Holzinger and Klaus-Martin Simonic, editors, *Information Quality in e-Health*, pages 673–679, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

Microsoft. Seeing AI. Turning the visual world into an audible experience. `https://www.microsoft.com/en-us/ai/seeing-ai`, 2020. Accessed: August 8, 2020.

BlindSquare. Pioneering accessible navigation – indoors and outdoors. MIPsoft, March 2020. `http://www.blindsquare.com/`.

Alicia Montarzino, Brian Robertson, Peter Aspinall, Ana Ambrecht, Cathy Findlay, Julian Hine, and Bal Dhillon. The impact of mobility and public transport on the independence of visually impaired people. *Visual Impairment Research*, 9(2-3):67–82, 2007.

Karin Müller, Christin Engel, Claudia Loitsch, Rainer Stiefelhagen, and Gerhard Weber. Traveling more independently: A study on the diverse needs and challenges of people with visual or mobility impairments in unfamiliar indoor environments. *ACM Transactions on Accessible Computing*, 15 (2), 2022. doi: 10.1145/3514255.

Masayuki Murata, Dragan Ahmetovic, Daisuke Sato, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. Smartphone-based indoor localization for blind navigation across building complexes. In *2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–10. IEEE, 2018.

Raghul Nagarajan, Sazali Yaacob, and G. Sainarayanan. Role of object identification in sonification system for visually impaired. In *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*, volume 2, pages 735–739, 2003. doi: 10.1109/TENCON.2003.1273276.

Eva-Maria Neumann. Egozentrische textuelle Beschreibungen von Gebäuden für Menschen mit Sehbeeinträchtigung. Master's thesis, Karlsruhe Institute of Technology, Department of Informatics, Karlsruhe, 2022.

New England ADA Center. ADA checklist for existing facilities. `https://adachecklist.org/doc/fullchecklist/ada-checklist.pdf`, 2016. Accessed: June 22, 2021.

John Nicholson. *Generation and analysis of verbal route directions for blind navigation.* PhD thesis, Utah State University, 2010. https://digitalcommons.usu.edu/etd/672/.

Hugo Nicolau, Tiago Guerreiro, and Joaquim A. Jorge. Designing guides for blind people. In *International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion*, New York, NY, USA, 2009a. ACM.

Hugo Nicolau, Joaquim Jorge, and Tiago Guerreiro. Blobby: How to guide a blind person. In *Extended Abstracts on Human Factors in Computing Systems (CHI EA)*, pages 3601–3606, New York, NY, USA, 2009b. ACM. doi: 10.1145/1520340.1520541.

Jakob Nielsen. *Usability engineering*. Elsevier, 1994.

Jakob Nielsen. Iterative design and testing of icon usability. https://www.nngroup.com/articles/icon-usability-1995-sun-microsystems-website/, 1995. Accessed: May 26, 2023.

Nielsen, Jakob. First Rule of Usability? Don't Listen to Users. https://www.nngroup.com/articles/first-rule-of-usability-dont-listen-to-users/, 2001. Accessed: April 12, 2023.

How Many Test Users in a Usability Study? Nielsen Norman Group, 2019. https://www.nngroup.com/articles/how-many-test-users/.

Sarah H. Norgate. Accessibility of urban spaces for visually impaired pedestrians. *Proceedings of the Institution of Civil Engineers - Municipal Engineer*, 165(4):231–237, 2012. doi: 10.1680/muen.12.00019.

Novartis. Novartis pharmaceuticals launches the first app for visually impaired people for use with the apple watch and other smart watches. https://www.novartis.com/news/media-releases/novartis-pharmaceuticals-launches-first-app-visually-impaired-people-use-apple-watch-and-other-smart-watches, 2015. Accessed: July 29, 2022.

Lazzus. Nuevo Sentido Tecnológico Realidad Aumentada S.L., January 2020. http://lazzus.com/.

OrCam Technologies Ltd. Orcam MyEye 2. https://www.orcam.com/en/myeye2/, 2020. Accessed: August 8, 2020.

Vera Zasũlich Pẽrez and Mauricio Santĩs Chaves. Haptic interfaces: kinesthetic vs. tactile systems. *Revista EIA*, 13(N.26):13–29, 2016.

Devi A. Paladugu, Hima Bindu Maguluri, Qiongjie Tian, and Baoxin Li. Automated description generation for indoor floor maps. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 211–212, New York, NY, USA, 2012. ACM. doi: 10.1145/2384916.2384958.

Jun Park and Subeh Chowdhury. Investigating the barriers in a typical journey by public transport users with disabilities. *Journal of Transport & Health*, 10:361–368, 2018. doi: 10.1016/j.jth.2018.05.008.

Amy Parker, Martin Swobodzinski, Julie Wright, Kyrsten Hansen, Becky Morton, and Elizabeth Schaller. Wayfinding tools for people with visual impairments in real-world settings: A literature review of recent studies. *Frontiers in Education*, 6, 2021. doi: 10.3389/feduc.2021.723816.

Gaëtan Parseihian, Charles Gondre, Mitsuko Aramaki, Sølvi Ystad, and Richard Kronland-Martinet. Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Transactions on Multimedia*, 18(4):674–686, 2016.

Clarissa Penfold, Nicky Cleghorn, Chris Creegan, Hayley Neil, and Stephen Webster. Travel behaviour, experiences and aspirations of disabled people. *London, England: National Centre for Social Research*, 2008.

Camille Peres, Christopher Frauenberger, Thomas Hermann, Tony Stockman, Louise Nickerson, John Neuhoff, Barbara Shinn-Cunningham, Virginia Best, and Derek Brock. *HCI Beyond the GUI: The Human Factors of Non-traditional Interfaces*, chapter 5. Auditory Interfaces, pages 147–196. 2008.

J. Eduardo Pérez, Myriam Arrue, Masatomo Kobayashi, Hironobu Takagi, and Chieko Asakawa. Assessment of semantic taxonomies for blind indoor navigation based on a shopping center use case. In *Proceedings of the 14th International Web for All Conference*, W4A, New York, NY, USA, 2017. Association for Computing Machinery. doi: 10.1145/3058555.3058575.

Darius Plikynas, Arunas Zvironas, Andrius Budrionis, and Marius Gudauskis. Indoor navigation systems for visually impaired persons: Mapping the features of existing technologies to user needs. *Sensors*, 20(3), 2020.

Denise Prescher, Jens Bornschein, Wiebke Köhlmann, and Gerhard Weber. Touching graphical applications: bimanual tactile interaction on the hyperbraille pin-matrix display. *Universal access in the information society*, 17:391–409, 2018.

Bettina Pressl and Manfred Wieser. A computer-based navigation system tailored to the needs of blind people. In Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Arthur I. Karshmer, editors, *Computers Helping People with Special Needs*, pages 1280–1286, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. Watchout: Obstacle sonification for people with visual impairment or blindness. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 402–413, New York, USA, 2019. Association for Computing Machinery. doi: 10.1145/3308561.3353779.

Project ATMaps. Specification of symbols used on audio-tactile maps for individuals with blindness (ATMAPS). `https://www.atmaps.eu/index.php/english`. Accessed: July 22, 2022.

Projekt m4guide. m4guide - mobile multi-modal mobility guide. 2. Datenbasis, Outdoor-Navigation. Technical report, Projekt m4guide: Unabhängig mobil, 2017. `https://web.archive.org/web/20170227131114/http://m4guide.de/files/Zusatzinfo%202%20Projekt%20m4guide.pdf`.

Pablo-Alejandro Quinones, Tammy Greene, Rayoung Yang, and Mark Newman. Supporting visually impaired navigation: A needs-finding study. In *Extended Abstracts on Human Factors in Computing Systems (CHI EA)*, pages 1645–1650, New York, NY, USA, 2011. Association for Computing Machinery. doi: 10.1145/1979742.1979822.

F. Ribeiro, D. Florêncio, P. A. Chou, and Z. Zhang. Auditory augmented reality: Object sonification for the visually impaired. In *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pages 319–324, 2012. doi: 10.1109/MMSP.2012.6343462.

Emma Pratt Richens, Gareth Ford Williams, Jamie Knight, and Rebecca (Ed.) Nancarrow. BBC Mobile accessibility guidelines v1.10. Technical Report WCAG 2.1, W3C, 2019.

RightHear. RightHear. What if signs could talk? `https://www.right-hear.com/`. Accessed: April 19, 2023.

RightHear. RightHear. Mapping & Installing the RightHear Solution. Technical Report DOE-SLC-6903-1, Ra'anana, Israel, 2018.

Sebastian Ritterbusch and Harald Kucharek. Robust and incremental pedestrian path network generation on openstreetmap for safe route finding. In Klaus Miesenberger and Georgios Kouroupetroglou, editors, *Computers Helping People with Special Needs*, pages 302–309, Cham, 2018. Springer International Publishing.

Sebastian Ritterbusch, Patrick Frank, and Vanessa Petrausch. Modular human-computer interaction platform on mobile devices for distributed development. In Klaus Miesenberger and Georgios Kouroupetroglou, editors, *Computers Helping People with Special Needs*, pages 75–78, Cham, 2018. Springer International Publishing.

David A. Ross and Bruce B. Blasch. Wearable interfaces for orientation and wayfinding. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies (ASSETS)*, pages 193–200, New York, NY, USA, 2000. ACM. doi: 10.1145/354324.354380.

Manaswi Saha, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. Closing the gap: Designing for the last-few-meters wayfinding problem for people with visual impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS)*, New York, NY, USA, 2019. Association for Computing Machinery.

G. Sainarayanan, Raghul Nagarajan, and Sazali Yaacob. Fuzzy image processing scheme for autonomous navigation of human blind. *Applied Soft Computing*, 7(1):257–264, 2007. doi: 10.1016/j.asoc.2005.06.005.

Jaime Sánchez and Natalia de la Torre. Autonomous navigation through the city for the blind. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 195–202, New York, NY, USA, 2010. ACM. doi: 10.1145/1878803.1878838.

Daisuke Sato, Uran Oh, Kakuya Naito, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. NavCog3: An Evaluation of a Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 270–279. Association for Computing Machinery, 2017.

Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M. Kitani, and Chieko Asakawa. Navcog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *ACM Transactions on Accessible Computing (TACCESS)*, 12(3), 2019a.

Daisuke Sato, Hironobu Takagi, and Chieko Asakawa. *Web Accessibility: A Foundation for Research*, chapter Wayfinding, pages 677–700. Springer London, London, 2019b. doi: 10.1007/978-1-4471-7440-0_34.

Boris Schauerte, Torsten Wörtwein, and Rainer Stiefelhagen. A web-based platform for interactive image sonificationn. In *Accessible Interaction for Visually Impaired People (AI4VIP)*, Stuttgart, Germany, 2015.

Joram Schito. Effizienzanalyse der akustischen Wahrnehmung einer Parameter Mapping Sonification eines digitalen Höhenmodells durch interaktive Datenexploration: Masterarbeit über die Verwendung von Sonifikation in der GIScience. Master's thesis, ETH Zurich, Geographisches Institut, 01 2012.

Eldon Schoop, James Smith, and Bjoern Hartmann. Hindsight: Enhancing spatial awareness by sonifying detected objects in real-time 360-degree video. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2018. Association for Computing Machinery. doi: 10.1145/3173574.3173717.

Seeing Eye GPS. Senderogroup, January 2020. `http://www.senderogroup.com/products/SeeingEyeGPS/index.html`.

Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Education Limited, global edition, 6th edition, 2018.

Walter C. S. S. Simões, Guido S. Machado, Andre M. A. Sales, Mateus M. de Lucena, Nasser Jazdi, and Vicente F. de Lucena. A review of technologies and techniques for indoor navigation systems for the visually impaired. *Sensors*, 20(14), 2020. doi: 10.3390/s20143935.

Ruby Spencer. 8 Ways to Reduce Cognitive Load in eLearning. `https://elearningindustry.com/8-ways-reduce-cognitive-load-elearning`, 2015.

Martin Spindler, Michael Weber, Denise Prescher, Mei Miao, Gerhard Weber, and Georgios Ioannidis. Translating floor plans into directions. In *International Conference on Computers for Handicapped Persons*, pages 59–66. Springer, 2012.

Julian Striegl, Julius Felchow, Claudia Loitsch, and Gerhard Weber. Accessible indoor orientation support by landmark-based navigation. In Margherita Antona and Constantine Stephanidis, editors, *Universal Access in Human-Computer Interaction*, HCII, pages 510–524, Cham, 2023. Springer Nature Switzerland.

Thomas Strothotte, Steffi Fritz, Rainer Michel, Andreas Raab, Helen Petrie, Valerie Johnson, Lars Reichert, and Axel Schalt. Development of dialogue systems for a mobility aid for blind people: Initial design and usability testing. In *Proceedings of the Second Annual ACM Conference on Assistive Technologies (ASSETS)*, pages 139–144, New York, NY, USA, 1996. Association for Computing Machinery. doi: 10.1145/228347.228369.

Tony Stubblebine. *Regular expression pocket reference*. O'Reilly Media, Inc., 2003.

Ingeborg Stude, Ellen Müller, Burkhard Lüdtke, Robert Niemann, and Annette Müller. Berlin - Design for all: Öffentlich zugängliche Gebäude. Technical report, Senatsverwaltung für Stadtentwicklung und Umwelt, Berlin, 2012.

Swiss Federation of the Blind. Myway lite. `https://itunes.apple.com/us/app/myway-lite/id494516234?mt=8`, 2018. Accessed: September 21, 2018.

Titus J. J. Tang and Wai Ho Li. An assistive eyewear prototype that interactively converts 3d object locations into spatial audio. In *Proceedings of the ACM International Symposium on Wearable Computers (ISWC)*, pages 119–126, New York, NY, USA, 2014. Association for Computing Machinery. doi: 10.1145/2634317.2634318.

Technical Committee ISO/TC 159, Ergonomics, Subcommittee SC 4, Ergonomics of human–system interaction. Ergonomics of human-system interaction - part 112: Principles for the presentation of information. Standard ISO 9241-112:2017, International Organization for Standardization and International Electrotechnical Commission, 2017.

Neela Thapar, Grace Warner, Mari-Lynn Drainoni, Steve Williams, Holly Ditchfield, Jane Wierbicky, and Shanker Nesathurai. A pilot study of functional access to public buildings and facilities for persons with impairments. *Disability and rehabilitation*, 26:280–9, 2004. doi: 10.1080/09638280310001649543.

Kay Tislar, Zackery Duford, Madeline Peabody Brittany Nelson, and Myounghoon Jeon. Examining the learnability of auditory displays: Music, earcons, spearcons, and lyricon. In *Proceedings of the 24th International Conference on Auditory Display (ICAD)*, pages 197–202, 2018. doi: 10.21785/icad2018.029.

Transition Technologies S.A. Seeing assistant move. `https://seeingassistant.tt.com.pl/move/`, 2018. Accessed: September 20, 2018.

Johannes Tröger, Sarah Schnebelt, and Jan Alexandersson. *Technological Trends in Improved Mobility of the Visually Impaired*, chapter Modelling the Creation of Verbal Indoor Route Descriptions for Visually Impaired Travellers, pages 355–377. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-030-16450-8_15.

TUD. Ergebnis aus der Anforderungsanalyse AP2 (TUD): Konzeptionelles Datenmodell Version 4. Technical report, Projekt MOBILITY: Mobile zugängliche Lagepläne, 2011.

Christoph Urbanietz, Gerald Enzner, Alexander Orth, Patrick Kwiatkowski, and Nils Pohl. A radar-based navigation assistance device with binaural sound interface for vision-impaired people. In *Proceedings of the 25th International Conference on Auditory Display (ICAD)*. Department of Computer and Information Sciences, Northumbria University, 2019. doi: 10.21785/icad2019.023.

UsabilityHub. Testing iconography. `https://usabilityhub.com/guides/testing-iconography`. Accessed: May 26, 2023.

Joey Van der Bie, Christina Jaschinski, and Somaya Ben Allouch. Sidewalk, a wayfinding message syntax for people with a visual impairment. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 609–611, New York, NY, USA, 2019. Association for Computing Machinery. doi: 10.1145/3308561.3354625.

Frank van Diggelen and Per Enge. The world's first gps mooc and worldwide laboratory using smartphones. In *Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation, ION GNSS+*, pages 361–369, 2015.

Antonio Vasilijevic, Kristian Jambrosic, and Zoran Vukic. Teleoperated path following and trajectory tracking of unmanned vehicles using spatial auditory guidance system. *Applied Acoustics*, 129: 72–85, 2018. doi: https://doi.org/10.1016/j.apacoust.2017.07.001.

Karen Vines, Chris Hughes, Laura Alexander, Carol Calvert, Chetz Colwell, Hilary Holmes, Claire Kotecki, Kaela Parks, and Victoria Pearson. Sonification of numerical data for education. *Open Learning: The Journal of Open, Distance and e-Learning*, 34(1):19–39, 2019. doi: 10.1080/02680513.2018.1553707.

Mark Vollrath, Kathrin Leske, Berhard Friedrich, and Steffen Axer. Innerstädtische mobilitätsunterstützung für blinde und sehbehinderte: Inmobs: Schlussbericht: Teilvorhaben technische universität braunschweig. *Technische Universität Braunschweig*, 2015.

Bruce Walker and Gregory Kramer. *Ecological Psychoacoustics*, chapter Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making, pages 150–175. 2004. doi: 10.1163/9780080477442_009.

Bruce Walker and Lisa Mauney. Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing (TACCESS)*, 2:12, 2010. doi: 10.1145/1714458.1714459.

Hélène Walle, Cyril De Runz, Barthélemy Serres, and Gilles Venturini. A survey on recent advances in ai and vision-based methods for helping and guiding visually impaired people. *Applied Sciences*, 12(5):2308, 2022.

Change the Definition of Blindness. WHO, June 2020. `https://www.who.int/blindness/Change%20the%20Definition%20of%20Blindness.pdf`.

Michele A. Williams, Amy Hurst, and Shaun K. Kane. "Pray before you step out": Describing personal and situational blind navigation behaviors. In *Proceedings of the 15th International ACM*

SIGACCESS Conference on Computers and Accessibility (ASSETS), New York, NY, USA, 2013. Association for Computing Machinery. doi: 10.1145/2513383.2513449.

Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. "Just let the cane hit it": How the blind and sighted see navigation differently. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS)*, pages 217–224, New York, NY, USA, 2014. Association for Computing Machinery. doi: 10.1145/2661334.2661380.

Michelle A. Williams, B Dubin, Clara Amaefule, L Nguyen, A Abdolrahmani, C Galbraith, A Hurst, and Shaun Kane Kane. Better supporting blind pedestrians and blind navigation technologies through accessible architecture. In *Designing around people*, pages 237–246. Springer, 2016.

J. Wilson, B. N. Walker, J. Lindsay, C. Cambias, and F. Dellaert. Swan: System for wearable audio navigation. In *11th IEEE International Symposium on Wearable Computers*, pages 91–98, 2007. doi: 10.1109/ISWC.2007.4373786.

C-J Yang, J-F Zhang, Y Chen, Y-M Dong, and Y Zhang. A review of exoskeleton-type systems and their key technologies. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 222(8):1599–1612, 2008. doi: 10.1243/09544062JMES936.

Jing Yang, Yves Frank, and Gábor Sörös. Hearing is believing: Synthesizing spatial audio from everyday objects to users. In *Proceedings of the 10th Augmented Human International Conference (AH)*, New York, NY, USA, 2019. Association for Computing Machinery. doi: 10.1145/3311823.3311872.

Rayoung Yang, Sangmi Park, Sonali R. Mishra, Zhenan Hong, Clint Newsom, Hyeon Joo, Erik Hofer, and Mark W. Newman. Supporting spatial awareness and independent wayfinding for pedestrians with visual impairments. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 27–34, New York, NY, USA, 2011. Association for Computing Machinery. doi: 10.1145/2049536.2049544.

Po-Yin Yen and Suzanne Bakken. A comparison of usability evaluation methods: Heuristic evaluation versus end-user think-aloud protocol – an example from a web-based communication tool for nurse scheduling. *AMIA Annual Symposium Proceedings*, pages 714–718, 2009.

Clint Zeagler. Where to wear it: Functional, technical, and social considerations in on-body location for wearable technology 20 years of designing for wearability. In *Proceedings of the ACM International Symposium on Wearable Computers (ISWC)*, pages 150–157, New York, NY, USA, 2017. Association for Computing Machinery. doi: 10.1145/3123021.3123042.

Limin Zeng, Denise Prescher, and Gerhard Weber. Exploration and avoidance of surrounding obstacles for the visually impaired. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, pages 111–118, New York, NY, USA, 2012. Association for Computing Machinery. doi: 10.1145/2384916.2384936.

Limin Zeng, Markus Simros, and Gerhard Weber. Camera-based mobile electronic travel aids support for cognitive mapping of unknown spaces. In *Proceedings of the 19th International Conference* *on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)*, New York, NY, USA, 2017. Association for Computing Machinery. doi: 10.1145/3098279.3098563.

Tim Ziemer and Holger Schultheis. Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks. *Journal on Multimodal User Interfaces*, 13:205–218, 11 2018.

Thomas Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A hand gesture interface device. In *ACM Sigchi Bulletin*, volume 17, pages 189–192, 1986. doi: 10.1145/30851.275628.