Digital Digits: Designing Assistive Finger Augmentation Devices

by

Roy Shilkrot

M.Sc., Tel-Aviv University (2011)
B.Sc., Tel-Aviv-Yaffo College (2007)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

Signature redacted

Author ....
Program in Media Arts and Sciences
August 14, 2015

Signature redacted

Certified by... ....
Pattie Maes
Professor of Media Arts and Sciences
Program in Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by.. ....
Pattie Maes
Academic Head
Program in Media Arts and Sciences

Digital Digits: Designing Assistive Finger Augmentation Devices

by

Roy Shilkrot

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on August 14, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Media Arts and Sciences

Abstract

Wearable computers are becoming a widespread reality. Driven by a quest for sensorial *ultrability* (ultra-ability) and control of our environment and bodies, we search for ever more intimate solutions to increase our innate physical capacities using technology. Finger-wearable devices for augmentation are nowadays part of the mainstream wearable fashion and research agenda, because of their uniquely convenient placement on the human body and proximity to the most expressive of limbs - the fingers.

This thesis proposes a consideration of finger augmenting devices as a new class of instruments, rather than an opportunistic approach for positioning sensors and actuators. Out of a comprehensive survey of the work on finger augmentation, I put forward a definition for finger augmentation, a classification framework, and design guidelines for creating new finger-worn devices. I present four designs of finger-augmenters, their technical underpinnings, evaluation methods and theoretical contributions.

Assistance is ubiquitous throughout the spectrum of technological benefit, advancing those with specific needs for recovery or rehabilitation, as well as those looking to go beyond human ability. This cross-cutting design principle for human-computer interfaces is uncontested yet underutilized. This thesis conceptualizes the Assistive Augmentation spectrum as a metaphor for the flexible interpretability of technology to simultaneously help many communities. The concrete prototypes I hereby present: EyeRing, FingerReader, Mobile-FingerReader and MusicReader, exemplify this idea and suggest an inclusive path of technology development.

Thesis Supervisor: Pattie Maes
Title: Professor of Media Arts and Sciences, Program in Media Arts and Sciences

Digital Digits: Designing Assistive Finger Augmentation Devices

by

Roy Shilkrot

The following people served as readers for this thesis:

Signature redacted

Thesis Reader ....

Joseph A. Paradiso
Associate Professor of Media Arts and Sciences
Program in Media Arts and Sciences

Signature redacted

Thesis Reader ....

Kenneth Perlin
Professor of Computer Science
New York University

Acknowledgements

I would like to thank Pattie Maes, my advisor, for letting me run free and explore strange ideas of a connection between humans and technology. When I first came to the Media Lab, I did not realize the gravity of luck I received when Pattie took me on a student. She deserves every bit of praise for the tireless support, trust, advice and understanding, however I am sure she is too humble to admit to any of these things. I thank my PhD committee, Joe Paradiso and Ken Perlin, for their support in the process of producing this thesis. A big thanks goes out to my former advisor, Danny Cohen-Or, beyond his trust and kindness, he never failed to encourage and dispense life and career advice, as well as spark my interest in computer graphics. And to my other mentors in MIT and beyond: Nick Montfort, Seth Teller, Ariel Shamir, Eyal Ofek and Ran Gal, thank you!

To my group members of Fluid Interfaces in the Media Lab. You never stopped amazing me with innovative ideas and endless inspiration to be a better innovator and researcher. Dear Suranga Nanayakkara, without whom I would never have found this incredible subject for research and the quest of assistive technology. Jochen Huber, an incredible HCI researcher, wonderful collaborator and a great person to be around. The most talented Amit Zoran, who inspired me to think about philosophy and life in a way no one ever did, and at the same time pushed me to the limits of my technical abilities. All of the Fluids: Natan Linder, Rony Kubat, Pol Pla, Marcelo Coelho, Aaron Zinman, Doug Fritz, Pranav Mistry, Sajid Sadi, Anette Von Kapri, Jürgen Steimle, Seth Hunter, Valentin Heun, Xavi Benavides, Judith Amores, Scott Greenwald, Sang Won Leigh, Niaja Farve, Rahul Namdev, Harshit Agarwal, Anirudh Sharma. To my Media Lab companions, without which this journey was too dull to bear: Karthik Dinakar, Sean Follmer, Dave Mellis, Nan-Wei Gong, Gershon Dublon, Javier Hernandez, Daniel Leithinger. Lastly, delightful Amna Carreiro, Linda Peterson, Keira Horowitz and Alexandra Kahn, for all the help and support through years of work.

My work was driven and supported by wonderful MIT undergraduate students: Sophia, Michael, Emma, Sabrine, Connie, Albert, Happy, Tal, Erik, Kwadwo, Yashod, Kerry, Kevin, and amazing M.Eng. students: Marcelo Polanco and Kelly Ran.

A sincere thanks goes out to all user studies participants. To VIBUG, an amazing group of people who are set on making the lives of visually impaired persons better using technology and are willing to try anything. I also thank my sponsors: Steelcase and NEC, for supporting my work at the Media Lab.

To my dearest family in Israel: Baruch, Asia, Liraz and Orit, you are truly my inspiration and the strength you give is my lifeline. To my "framily" here in greater Boston: Yair, Sarit, Gilad, Manu, Yonatan, Noa (and Ayala), Amir, Michal (and Eitan), Guy, Talia, Yadid, Ifaat (and Noga, Ori), Eyal "Person", Shahar, Tamar, Nadav, Ami, thank you for being a surrogate family, thank for all the breakfasts, brunches, lunches, dinners, movies nights and frolic, you will never be forgotten.

And finally, my beautiful and beloved little family, Inbal and Roni, you have my endless uncontested gratitude and love for enduring with me during these testing times, lifting me up while keeping me grounded. I could not have received a greater gift than having you in my life. A thousand times: Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The days of the first wristwatch marked the beginning of our journey into the realm of wearable computers. Perhaps due to the intimate personal connection, the feeling of possession, or convenient direct manipulation, it seems nothing can now satisfy the voracious need for bodily augmentation technology. To explain this, Billinghurst draws a line between timekeeping of the ancient world and nowadays computing: the ever pressing move of technology into the towns, homes, hands and finally bodies of people [21]. Within this booming agenda are *finger wearable devices.* A fascinating clash of age-old tradition with cutting-edge technology is embodied in these finger worn artifacts that epitomize intimacy and innovation, and is the topic of this dissertation.

Similar to Isssac Asimov's rules for robots, Steve Mann postulated a number of ground rules for the present day cyborg - the augmented human. Mann's core ideas are of a non-restrictive, always-on system that is easily controllable but at the same time supports expressiveness and a means of communication [130]. These traits are easily identifiable in a finger-worn device. The most simple and inherent gesture of our bodies is the *Pointing* gesture (exhibited in children of 8 months), where we raise a finger towards a thing, a person, an immediate or faraway place, or even an abstract idea (like the "heavens") as a way of communication. This highest degree of bodily control is naturally paired with the body's most sensitive sensors - the fingertips. Bearing the densest network of tactile, thermal and pain sensors as well as tendon, blood vessel and bone, the fingers are highly represented in the brain's somatosensory and motor cortices, making them the perfect candidate for non-restrictive controllable augmentation.

In spite of these appealing traits, finger augmentation hasn't received the lion's share of attention from wearable computing researchers, and only in very recent years a clear research agenda started to emerge. My thesis offers, for the first time, a consideration of *Finger Augmenting Devices* as a new class of wearable computers - a standalone vector of development. In the coming chapters I will define, categorize, exemplify and evaluate Finger Augmenting devices, through four projects of my own creation and a broad examination of other work in this new exciting field.

1.1 Motivation

> "Reality has always been too small for human imagination."
> – Brenda Laurel, PhD Dissertation

Why research finger augmentation? The personal inspiration for picking this research subject is illusive but clearly visible: narratives of augmentation. From the dawn of mankind, stories and legends tell of hand wielded objects whose capabilities extend beyond their native performance. Moses' staff, Thor's hammer, King Arthur's Excalibur, and Sauron's rings are all archetypes of manual devices that empower their possessors. These often have intelligence and purpose of their own and perform in a unique, extended quality that far surpasses the standard object of their kind. Recently, these legends materialize via digital technologies that enable augmentation of hand worn devices, creating new computational extensions of the human body.

With or without impairments, people find themselves at the edge of sensorial and manual capability and seek assistive or enhancing devices. The overarching topic of this thesis is centered on the design and development of assistive technology, user interfaces and interactions that seamlessly integrate with a user's mind, body and behavior, providing an enhanced perception and motor control. I call this "Assistive Augmentation".

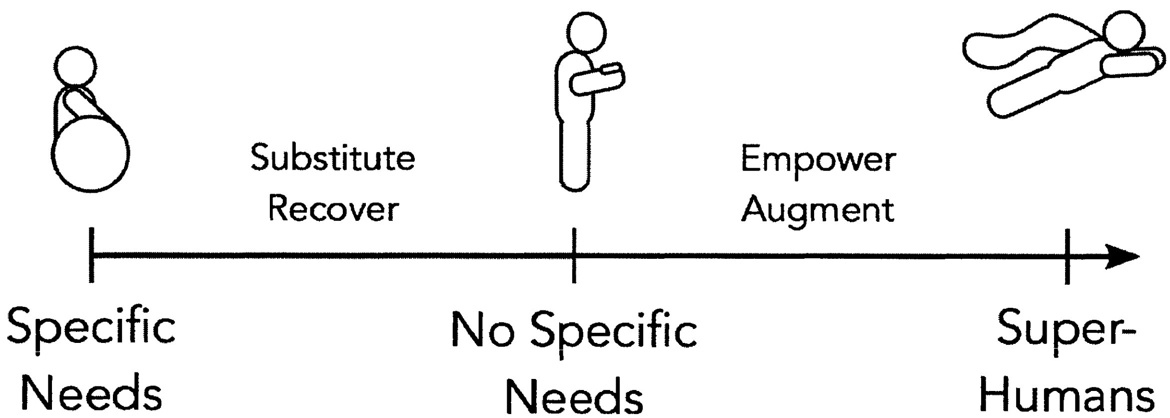


Figure 1-1: Assistive Augmentation continuum, a guideline for developing cross-domain assistive technology.

Assistive augmentation finds its application in a variety of contexts, for example in providing a scaffold for people when they feel their innate senses are inadequate or to support development of desired skillsets. I wish to put sensorial ability and disability on a continuum of usability (see Fig.1-1), rather than treat one or the other extreme as the focus. I therefore follow the design rationale of Rekimoto [180] stating technology should be socially acceptable, work coherently for people with and without impairments alike, and support independent and portable interaction. The latter requirement challenges both user interface and interaction design in particular, as Jones and Marsden point out: "the test comes when it [the device] is deployed in the complex, messy world of real situations [...] when it has to be used in the wild, as it were, in tandem with the world around it, the usability can break down quickly" (cf. [86], p. 51).

1.2 Assistive Augmentation

> "there isn't one preferred interpretation of a system but multiple. [...] technologies are *interpretively flexible*, i.e., lend themselves to different interpretations besides those intended by their developers"
> – Yvonne Rogers, HCI Theory

Assistive Augmentation is a framework I used for organizing my own work, but may also offer other researchers a basis to build on in their work. According to Yvonne Rogers, a conceptual framework in the field of HCI, as opposed to a *paradigm*, *theory* or a *model*, is a set of core concepts, questions and principles to consider when designing novel user interfaces ([187] p.5).

As mentioned earlier, the core concept of Assistive Augmentation is that assistive technology can most times apply outside the intended targeted assistance modality, therefore it creates a continuum of usability rather than hotspots or extremes. For example, if we consider depth cameras as a technology, their conception was in the robotics domain, however recently they moved to the computer interaction domain (e.g. in the Kinect), and already there is a myriad of examples of how they can revolutionize the area of assistive technology for persons with visual impairments [38]. Assistive Augmentation hints that practically any human-facing technology is in fact assistive technology, and the level of assistance for a certain audience depends on the primary intention of the creator, but it need not be so. The notion of the flexibility of technology to be applied to uses other than the intention of the creator is already an existing discourse in the HCI community ([187] p.71), in particular drawing lines between technology for people with visual impairments and people with dyslexia [28].

Assistive Augmentation is descriptive, prescriptive, generative and ethnographic ([187] p.16) framework. In the coming chapters of this document the reader can find the following elements:

- a terminology for researchers to speak about cross-boundary assistive wearable technology;
- guidelines for technology designers to contemplate impact on, or at least consideration of, using their proposed technology in additional application domains;
- help for creators to generate novel uses for a technology they developed with a specific use in mind;
- accounts of laboratory and in-the-field user studies with the target-audience that test validity of the proposed interactions.

1.3 Contributions

This dissertation contributes concrete exemplars of finger augmentation devices, and a broad defining discourse on the context of their materialization. The document also presents the design rationale behind creating such devices, their internal technical constructs, and their evaluation methods and results.

The thesis makes the following technical contributions:

1. Computer vision algorithms for finger-perspective imaging

(a) *Detecting the fingertip in a close-up video stream* using skin-detection and an efficient coarse-to-fine tracking method.

(b) *Text extraction from close-up video imaging*: un-distortion of the image, detecting lines of text, segmentation and extraction of words, tracking text lines and words.

(c) *Musical stave notation analysis from video imaging*: a first-in-kind, real time close-up video analysis of musical notation including staff line detection and removal, symbol and pitch classification, symbol and staff line tracking.

(d) *Imaging-based applications for the finger pointing gesture*: locating the pointing direction on a screen using visual features, detecting barcode features and extracting price tags.

2. *Evaluation and data-analysis methods for measuring finger-based reading efficacy.* I present three global quantitative metrics of efficiency of reading text in terms of consecutive words read, total words read, and tracking lines of text, as well as questionnaires that gauge the perceived efficiency qulitatively.

In addition, the thesis contributes detailed accounts of studies performed with the target audience of the respective devices, as well as a broad survey of the field of work. The following are the concrete contributions:

1. A broad overview of the field of finger augmentation citing over 150 works from the last 50 years.
2. A design guideline for creating assistive finger augmenting devices, considering aspects in both technical and nontechnical domains.
3. Laboratory evaluations of four finger-augmenting devices with qualitative and quantitative methods.

1.4 Document outline

The following is a description of the chapters in this document. My overall intention in this thesis was to tell the story of my work in chronological order, with one exception, the in-depth research of finger augmenting devices (given in Chapter 2) was performed after the initial work on the EyeRing (given in Chapter 3). The main body of the thesis contains the descriptions of the four major projects, including their respective evaluation work, rather than having a central evaluation chapter. I hope this construct will allow the reader to also perceive, between the lines and words, the evolution of my personal attitude towards finger augmentation that gradually gravitated towards assistive technology and working with people with visual impairments. The conclusion section will provide the readers with a recapitulation of the main ideas, but also a chance to hear of some reflexive discourse on the work.

1.4.1 Chapter 2: Background

This chapter covers the theoretical grounding behind my work on finger augmenting devices, presenting the concepts of Embodiment and Ubiquitous computing and their philosophical underpinnings. I also cover the relation of human augmentation, and within it finger augmentation, to Augmented Reality. The background chapter also contains a wide survey of finger augmentation work through the last century as presented in the academic and industrial worlds. I present a definition for Finger Augmenting Devices, and a breakdown of the concept to the elements of augmentation: the sensors, the actuators and the form factor. Each element is further scrutinized to offer a mapping of the design space, before finally presenting a guideline for finger augmentation researchers to consider when working in this field.

1.4.2 Chapter 3: The EyeRing

The EyeRing was the first finger-worn device I developed as part of this research endeavor, which presents the initial exploration of this design space and motivating ideas. The device is a finger-worn camera taking stills photos (unlike a video stream) that meant to be used as a personal assistant in a number of applications: assistance for people with a visual impairment in shopping, office-based applications and even reading music. The EyeRing is wireless, and pairs via BlueTooth to a mobile device where the computation takes place. This chapter discusses the hardware, different software elements, applications, and an evaluation of the EyeRing in comparison to using a regular mobile phone.

1.4.3 Chapter 4: The FingerReader

The FingerReader is a text-reading finger wearable video camera aiming to provide people with a visual impairment an easier access to printed text. The small camera is mounted on a finger worn construct, which also contains multiple vibration motors. The computer analyses the video image and produces audio and tactile feedback to the users, guiding their scanning process and speaking out the text out loud. The chapter covers the hardware and software components, including the computer vision algorithms, and an evaluation with visually impaired users of their usage of the device in different levels.

1.4.4 Chapter 5: The Mobile-FingerReader

The Mobile-FingerReader is a wearable smartphone camera peripheral that is a continuation of the work on the FingerReader device. The Mobile-FingerReader is considerably smaller and lighter in form, contains different camera hardware, does not have vibration capabilities, and strictly works with a smartphone device. It presents a more realistic usage scenario than the original FingerReader, as mobility and usage outside the home are identified as important for visually impaired persons in the context of reading text. The chapter describes the innovation of the device from the FingerReader, in terms of hardware and software, and the additional components designed to make it work with an Android smartphone. The chapter also presents a larger scale user study with visually impaired persons, which also contains a quantitative component.

1.4.5 Chapter 6: The MusicReader

The MusicReader is a finger wearable camera targeted at assisting musicians with a visual impairment in reading printed music sheets. Current solutions for such musicians are far from practical in situations such as orchestral play, classroom work and spontaneous band play. The MusicReader borrows some ideas from the FingerReader in terms of scanning printed material, but presents a different modality of audio feedback that suits the specific needs of a music reading application. The chapter describes in length the computer vision framework developed in order to support this novel application, the feedback modality to the user, and an evaluation with blind musicians from Berklee College of Music.

1.4.6 Chapter 7: Conclusion

The final chapter of the thesis presents a reiteration of the ideas around finger augmentation and the central contributions from the design explorations. It also speaks to the possible future of finger augmentation, if the development momentum is maintained or declines.

Chapter 2

Background

Finger Augmentation is a subfield of *Wearable Computing* – a rising area of academic research as well as product development. Two key theoretical concepts in Human Computer Interface (HCI) that relate to wearable computing are *Ubiquitous Computing* and *Embodiment*, which flesh out the connection between human perception of the physical world and the affordances of objects within it. This chapter begins with a note on these theoretical concepts that frames future discussions throughout the thesis around finger wearable computer interface. Thereafter the readers will find a broad overview of finger augmentation technology, its roots in history and a hierarchical analysis framework of its subelements. Lastly, the chapter offers a discussion around the design of finger augmenting devices and their potential advantages in approaching various challenges in HCI.

2.1 Theoretical and Philosophical Grounding

> "The future of interaction lies not in the interface 'disappearing', but rather in the interface becoming even more visible"

– Paul Dourish, Where the Action Is

My work draws on lessons from the Third Paradigm of HCI theory [65], in particular that of Ubiquitous Computing and of Embodiment, both products of the HCI theorists of the 1990s. An especially interesting hypothesis, set forth by Paul Dourish, concerns the relation of interactive technology and human perception of the world. Dourish opens his book "Where the Action Is" by stating how in the dawn of modern computing, when computers were scarce and incredibly expensive, the computer's time was far more valuable than the user's time, and how this formed the concept of making the job easier for the computer and harder for the user [41]. Whereas today with contemporary computers, the balance should be different – computers must be highly responsive to the user's actions and needs.

The core of Dourish's ideas is embodiment, which he describes as "being manifest in as part of the world", a well established theory in the area of phenomenology, developed by Edmnud Husserl and Martin Heidegger in the first half of the 20th century. Interaction is an embodied action, meaning it moved from being an idea in one's mind to an experience in one's hand. An obvious connection of assistive augmentation to embodiment is the fact that sensorial augmentation, as is perception, is embedded in the real environment. In many cases the augmentation is mediated, as in the case of screens and cameras, but the constant central element is an embodied augmentation of a human sensory experience.

Dourish developed the idea of the importance of the human body and environment to computer interfaces stating that every activity is situated in and derives meaning from a particular environment and a moment in time, in addition to that, human understanding of the world stems from the ability to act within it and upon it, rather than only observe and contemplate it.

Ubiquitous Computing (UbiComp) is a concept of the future of HCI set forth by Mark Weiser [247] and Pierre Wellner [248]. Weiser defines UbiComp as an embedding of computers in everyday objects and the environment itself: walls, notepads, pens and even doors. Weiser's goal in UbiComp was to make technology invisible, and he also claims that indeed in Virtual Reality the computer is effectively invisible [246]. Nicola Liberati, a contemporary phenomenologist of Augmented Reality (AR) and UbiComp, imparts a central guideline to the creation of "good AR" technology: augmented objects should be as perceptually real as physical ones so that the technology itself would seem invisible. While UbiComp is not an enhancement of the *subject* like AR but of the *objects*, it shares the same goals as AR of exposing a clear affordance to the user [122].

These theorizations guided my work towards devices that offer augmentations directly of the hand of the user. For one, the computerized elements augment a well-practiced deictic gesture of pointing a finger. An important component of embodied interaction, according to my interpretation, is that the user must always remain in manual control of the action, to maintain an integration of the mind, hand and computer.

2.2 Finger Augmenting Devices: A Comprehensive Survey

Wearable ubiquitous computing is no longer a dream–it is the reality we live in. It has grown from a niche field of academic research into a multi-billion dollar industry, and a booming scholarly endeavor. The advent of wearable computers gave rise to *Finger Augmentation*, an up-and-coming domain of devices worn primarily on a finger to add sensing and feedback, and allow a new kind of manual interaction with the world. New *Finger-Augmenting Devices* (henceforth: FADs) appear annually in major academic venues of the HCI community (see Fig.2-1), on the consumer market as new products,

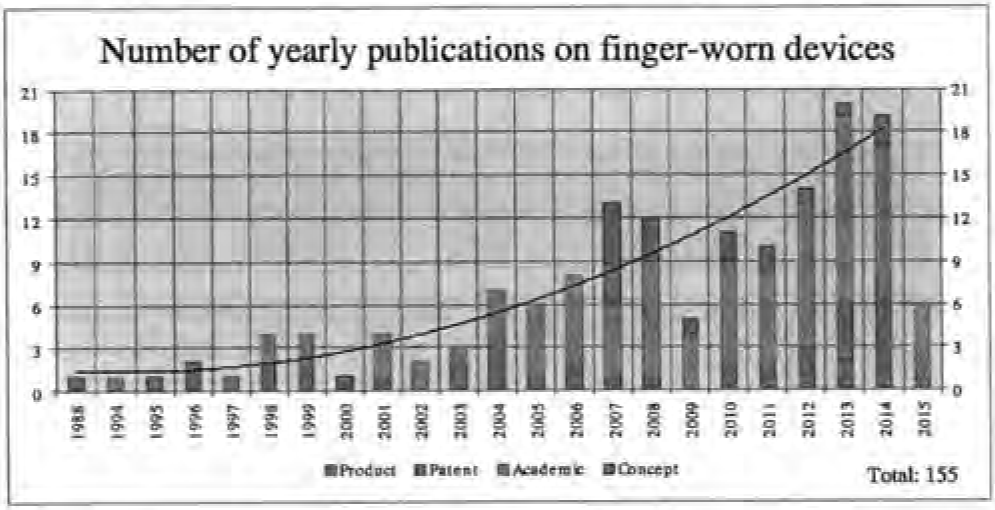


Figure 2-1: The yearly account of publications on FADs suggests a growing trend. Note: (1) we do not yet have a full tally of the works published beyond 2015, and (2) in the years 1916-1987 there were 4 publications that are not visualized here.

and the prominent news and popular media outlets. The demand for these types of devices is increasing, which is the reason we set upon the undertaking of surveying, documenting and defining the field.

Finger augmentation seeks to add two additional abilities to the innate human finger ones: (1) to sense (input) beyond what the ordinary human finger senses (e.g. image, tactile, thermal) and (2) to provide (output) information to the wearer, and (3) control or output information via the finger to an external object. Such devices leverage on the finger's direct interaction with proximal surfaces, the inherent focus of attention derived from pointing and touching, and building on the dream of the extended hand's reach into virtual and distal worlds. Recognizing the potential of enhancing the finger with additional I/O capabilities, researchers and inventors suggested a large number of ways to attach sensors and actuators to the finger. Readily available finger-augmenting consumer products already rely on inertial sensing with accelerometers, gyroscopes and magnetometers, and interpret their signals to recognize gestures. However, other sensing modalities such as attaching miniature cameras are on the rise.

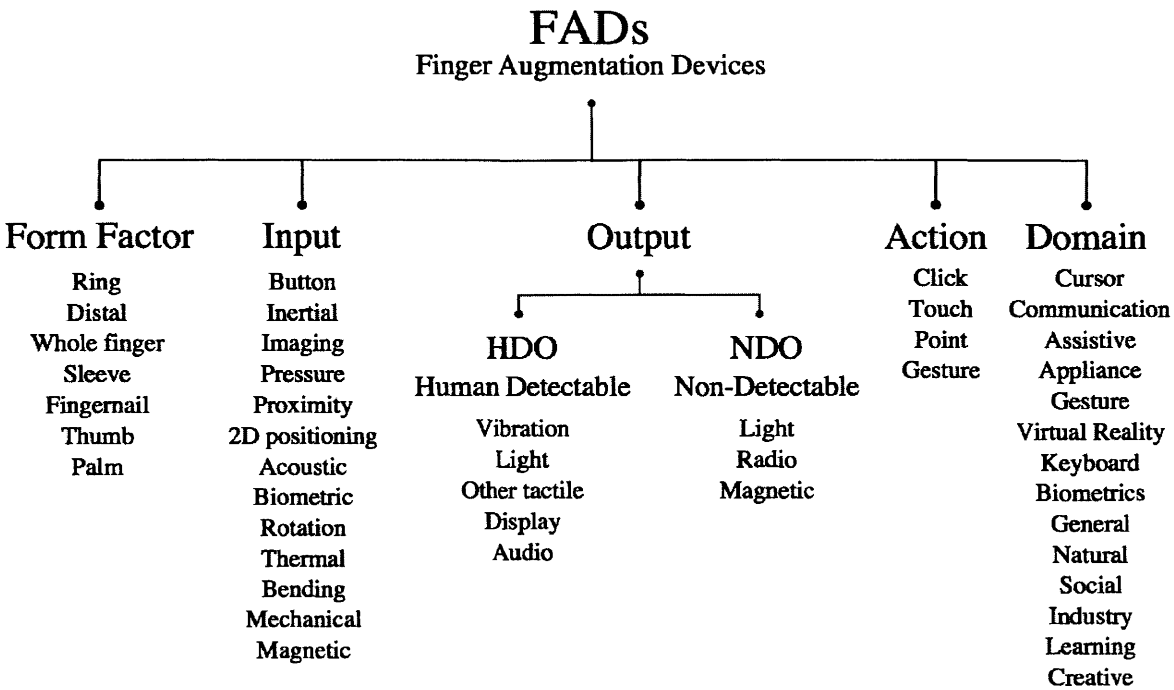


Figure 2-2: Classification hierarchy for FADs.

FADs come in myriad of shapes and forms, targeted at multiple audiences and applications. These FADs embed a wide range of sensors, power sources, wireless communication modules and actuators into very small form factors. The major application domain for these devices is keyboard and mouse input for personal devices, however applications in the medical, assistive and industrial domains are also very prominent. From controlling a cursor to preventing injury, each application domain drives the embedded components and interaction modalities. We created a classification based on the following categories rising from the works: input and output modalities; application domain; form factor and location on the finger; interaction scheme; and wireless capabilities (see Figure 2-2). While input, output, form factor, wireless capabilities and application domain categories compare the functions or the intended use of the devices, the interaction scheme category suggests a classification around where the interaction takes place: on the device itself, on a proximal surface, external (e.g. a remote object), etc.

The goal of this work is to provide an encompassing survey of the existing attempts at finger augmenting devices. Initially we provide our definition for such devices, separating them for example from smart glove interfaces or from finger-protectors of sorts. The discourse will center around the overarching range of technologies rising from the whole body of work rather than focus on a specific implementation, as each instance generally has different targets, intentions and evaluation methods. Previous surveys of finger-worn devices in standalone work or part of a publication on a specific implementation [153] cover specific subsets of the field: Rissanen et al. surveyed 20 works focusing on rings [185], and Nanayakkara et al. surveyed 15 instances with a broader viewpoint. We therefore believe our own work (surveying more than 140 instances) presents the most comprehensive, methodical and up-to-date overview of the entire field.

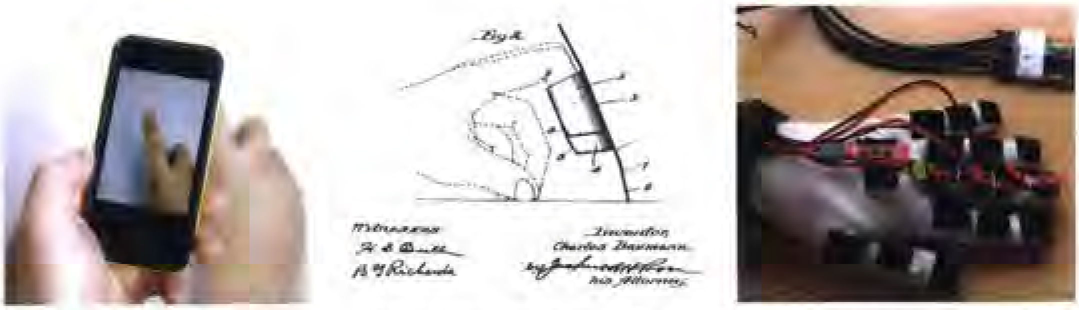
We begin with formative postulations on the boundaries of the field, contributing a working definition for a Finger Augmenting Device. A definition aids in separating other types of hand-worn devices, as well as makes ready a terminology to discuss FADs, which was not suggested to date. The second contribution is a classification hierarchy of categories and sub-parameters that is used to organize the surveyed literature and serve as organizational framework for future surveyance. We also contribute a set of guidelines to scaffold future research into FADs arising from the surveyed body of work. To make the knowledge more accessible, the text contains information boxes with a concise description of the merits and demerits of various approaches, which serve as bite-size pieces of advice for designers to consider. Finally, tables showing the actual classification of the work can be found in the appendix.

2.2.1 Definition of Finger Augmenting Devices

Hand-worn and hand-held augmentation devices are an incredibly large engineering and research endeavor of many years. Within this domain, finger-worn devices are a relatively new vector of investigation that was until recently inseparable from the larger agenda. Thus in order to create a boundary for the body of work on FADs, it is helpful to create a definition. As a trivial exercise in finding a good boundary for FADs, consider the regular computer mouse or a touch screen. These can be thought of as FADs since they gives our fingers abilities to move cursors on screens and perform click, drag, zoom and many other operations un-instrumented fingers cannot perform. However under this possible broad definition we should include just about any computer interface that involves a finger's touch, which accounts for most computer interfaces in existence. Since wearable computers become more popular and some also promise to provide special powers to our fingers, we tried to create a stringent rather than lenient definition for FADs using inherent qualities of the device itself rather than only its supposed function for our fingers. As a counter example, Digits [98], the OrCam glasses [162] or Nailsense [77] (see Fig.2-3a) also intend to augment the finger with capabilities, but they do not instrument the finger itself rather the glasses or mobile device.

We define finger-augmenting devices as *finger-worn devices with an additional augmentation other than their form, that provide a supplemental capability for one or more fingers using the finger itself as a central element.* Perhaps the hardest task in creating this definition was to separate the immense body of work on smart gloves [266], as they are also, to a degree, FADs. This distinction is nevertheless possible to make, for example in the work of [254] where the glove itself plays a central element in the sensing, whereas in [73] the glove only serves to provide a convenient mount for the researchers intending to augment the fingers (see Fig.2-3c).

We include *"additional augmentation other than their form"* in the definition for a FAD since some finger-worn objects do not provide a function beyond the affordances of their physical form. The following are examples of non-active finger augmentations that only provide a function via their form factor: a finger-worn stylus pen [219], a



(a) NailSense [77] (b) Finger Protector [30] (c) Hand pose sensor [73]

Figure 2-3: (a) and (b) are not considered FADs based on our definition, since they do not augment the finger itself (a) or have no augmentation other than their form factor (b). On the other hand, (c) is in fact a FAD since the glove is only used for conveniently placing sensors on the fingers. Images (a) and (c) are courtesy of their respective authors.

finger-worn painter's palette [11], a basketball training instrument [59] or even a self-defense device [101]. While these do provide an enhancement of the finger's inherent function, they do so only by a static supplement.

2.2.2 History of Finger Augmentation

The existence of finger wearables goes at least as far back as documented history. Finger rings in particular carried symbolic and mythical meaning throughout the ages of humankind, up to our own days [188]. In ancient times, ring devices were used to represent power (e.g. signet and seal rings of rulers), amuletic protection from evil spirits and bearing magical forces, while in more recent history used as ornaments and objects of remembrance and bond (e.g. an engagement ring). The narratives of ancient times, the tale of Prometheous' memento ring from Jupiter of the Greek mythology for example, still echo through our modern day media, where one would often see film and television scenes of couples buying and exchanging diamond wedding rings [137] (although the original symbolism may have been replaced by advertising [24]).

Finger wearables are intrinsically intertwined in our culture, however only "recent times" show functional usage for them beyond symbolism. Sewing thimbles (that date back to 206BC [2]) are an example of an ancient utilitarian finger augmenter, however the more recent abacus-ring (from 17th century China) and the document-sealing rings of the middle ages (10th - 15th centuries) are also of very practical usage, beyond being an emblem of status [137]. Even more recently, with the advent of the industrializtion age, finger wearable devices started to take much more of a practical role, usually as protectors such as the finger protector for safe kitchen knife operation [30] (see Fig.2-3b), or a device to assist in holding a writing pen [87].

Evidence form the early days of FADs, devices that operate beyond the factor of their form, are hard to pin down, however in 1916 and 1918 two patents were filed detailing a finger wearable electrical switch for easily operating the lights in a car [155; 63]. In 1965 a finger wearable switch to operate a sewing machine was patented [192], and since then FADs started to branch out into other input modalities such as a microphone [149] or a cursor-controlling pad [119].

2.2.3 Classification of FADs

In preparation for this survey, we collected works from academic publications, registered patents, currently available consumer products, and concept design works. Our collection includes 87 academic publications from conference proceedings, journals and theses, 29 patents, 16 consumer products and 16 design concepts of FADs. We also surveyed 23 other pieces that do not fit our definition for a FAD, nevertheless they are all relevant to the discussion. Pieces were collected using a number of methods: systematic search through conference proceedings for relevant work (in particular ACM CHI, UIST, TEI, UbiComp, ISWC), keyword search through academic and patent publication index engines, hierarchical citation-tree backtracking from existing publications, and lastly general queries in a web search engine.

The obvious advantage of an academic publication is that it presents the technical details in a clear fashion, however it often reflects a research-in-progress rather than a ready-for-market product. It was therefore important to seek out products to complete the picture, as these mostly represent a mature state of work. Patents present units of knowledge that exhibit enough market potential so they needed to protected, however they are often not in a high stage of maturity. Design concepts that are freely published online add a valuable aspect of *wishful engineering* and a futuristic outlook that brings out the needs and desires from FADs. This range spans the extent of the current research, implementation and ideation on FADs.

Our classification considers the following dimensions: form factor, input modality, output modality, the device's action and the application domain. The form factor, input and output modalities prescribe the physical affordances of a FAD, as they determine how it could be appropriated to interface with the user's hand. These categories were quantized in a combined inductive and deductive process, where both preconceived and emergent sub-categories were used to parametrize the surveyed work (see Figure 2-2). Classifying the intended action for a FAD was first indicated in [153] and further developed here to suggest that the inherent action of fingers and hands supports 4 levels of FAD interaction: touching a surface, gesturing in the air, touching the finger-worn device and pointing at a referent. Lastly we examine the rich world of applications for FADs, as they evidently cluster together to solve or enhance manual operations in a given field.

2.2.3.1 Form Factors

To support different interactions FADs are positioned and formed in many different ways. They attach to any of the finger sections: proximal, medial and distal, some cover more than one section to form sleeves and others cover the whole finger and even the palm and fingernail. We identify seven generic form factors used in FADs, which

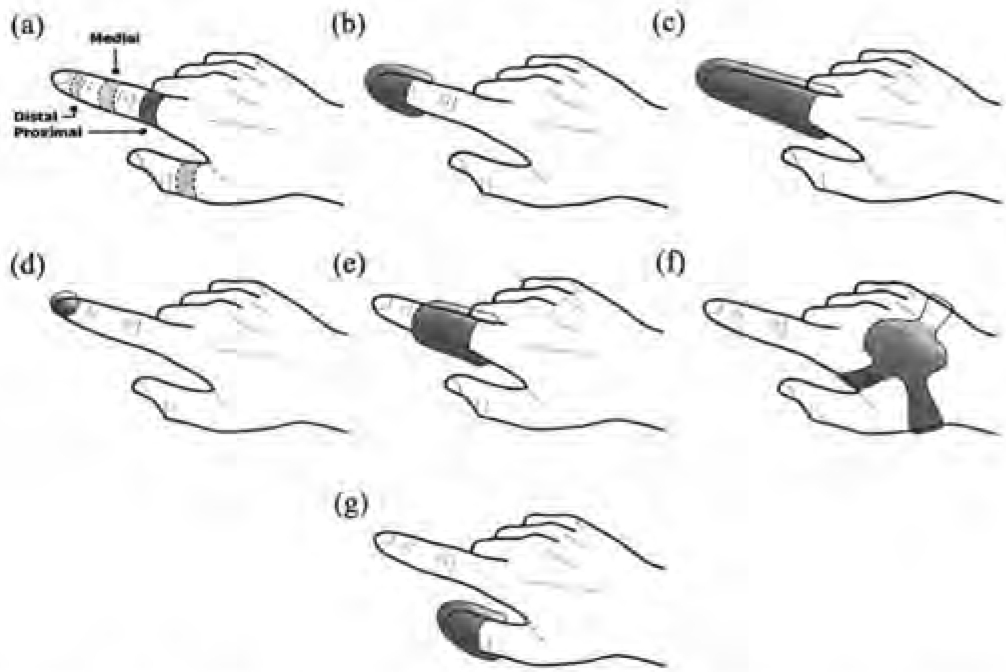


Figure 2-4: FAD form factors. (a) rings, (b) distal addendum, (c) whole finger addendum (d) fingernail addendum, (e) sleeve, (f) palm component that accompanies a FAD, (g) thumb addendum.

are also illustrated in Figure 2-4: rings, distal addendums, whole-finger addendums, fingernail addendums, sleeves, thumb addendums, and finally components mounted on the palm that support the FADs. Another varying parameter is the number of devices used in a single FAD, sometimes many rings are used on multiple fingers at the same time (such as in the case of chording keyboards [48; 70; 35; 12]) or a finger-thumb pair [119; 33], however we found that in more than 80% of the work there is only a single wearable device.

The most prominent form factor is the ring (see Fig. 2-4 (a)), which is considered to be the most acceptable and practical long-term wearable device due to the long history of finger-rings as jewelry. Moreover, rings present the least cumbersome form factor that leaves the hands free to grasp and interact with screens and other devices. Most of the rings are worn, according to traditional custom, on the proximal phalanx section of the finger, however unique cases show distal-rings for biometric purposes

[181], medial-rings for cursor control or pointing [214; 71] and medial-rings for gestural interaction [83; 105]. In the early days of FADs rings were strongly dominant, later to be surpassed in popularity by distal addendums and whole-finger addendums, and their applications gravitated towards input devices to augment or replace the mouse and keyboard. In recent incarnations ring FADs are used more and more as gestural and cursor control input devices.

Form factor: Rings

Pros: Socially acceptable form, with a rich cross-cultural tradition and narrative. Easily accessible for the thumb, since usually a ring rests around the proximal phalanx. Small in size, and most times easy to take off and wear.

Cons: Difficult to develop for such a small form, and using some input and output modalities is not feasible. There are social implications and preconceptions as to what is acceptable in/on a ring.

Main applications: communication, computer input.

Distal addendums, devices that attach to the end of the finger and sometimes cover it (see Fig. 2-4 (b)), are an overall runner-up to the predominant rings, however they distinctly differ from them in the applications they offer. Some applications are practically unique to distal addendums, such as interfaces for virtual and augmented reality [103; 170; 34]. Since distal addendums are located at the fingertips in an area of heightened sensing, as discussed in the last section, they are prime candidates to create output interfaces, and indeed over 75% of distal addendum FADs we covered pack some sort of output modality: tactile [104], vibration [263], light [257] or others.

Form factor: Distal addendums

Pros: Can output information via the sense of touch. Proximal to the touch surface, and allows for sensing it.

Cons: Obstructs the inherent sense of touch, and adds unnatural weight on the fingertip.

Main applications: virtual reality.

Whole-finger and sleeve-addendum devices cover bigger portions of the finger, and this real-estate allowed creators to explore a more encompassing monitoring of the finger (such as its bending [239; 120; 67]) as well as incorporate much more interaction elements, for example buttons [189; 94; 168] and high speed and accuracy tracking [174].

Form factor: Whole finger and Sleeve

Pros: Have much more room for input/output elements. Enable sensing bending, and easily lend to understanding the direction of pointing.

Cons: Big and cumbersome to wear and remove. May obstruct the natural motoric affordances of the fingers.

Main applications: computer input.

A new up-and-coming form factor for finger augmentation is that of nail addendum. In spite of the appealing characteristics of the nail as a bed for adding input and output (see the following information box), there were not many cases reported. Prince was the first to postulate the usage of nail augmentation to enhance interaction with computers [171], which was reiterated years later by [105]. Asada and Mascaro looked into understanding how pressure affects the fingernail's appearance to allow for the creation of virtual buttons [132; 133]. Others experimented with different input modalities mounted on the fingernail, such as a photo detector [6], a pressure sensor [91] (see Figure 2-5b) and a magnetometer [121; 29], while others mounted a vibrator motor [234] (see Figure 2-5c) and an RFID chip [240].

(a) Nail Display [229] (b) NailO [91] (c) SmartNail [234]

Figure 2-5: Finger-nail augmenting devices: (a) a miniature display added to the thumb, (b) a touch-senstive pad for 2D cursor control, and (c) a radio controlled vibration motor.

Form factor: Fingernail addendum

Pros: Fingernails are underutilized augmentation real estate. Very close to the touching surface, but don't obstruct the fingertip pad and the natural sense of touch. Allow using adhesives, and there are no nerve endings that allows for example "tapping" with the nail.

Cons: Wearing something on the fingernail carries a social meaning. Slightly inaccessible for the other fingers or thumb of the same hand. Added weight on the tip of the finger may be uncomfortable. Difficult form to design for, although the thumb nail is often larger.

Main applications: computer input, assistive technology.

2.2.3.2 Embedded Input Modalities

According to our statistics most FADs are input devices, where 119 of the 160 surveyed works reported having at least one kind of input modali. Our classification of FADs revealed a wealth of commonly used input modalities, each making use of a different element of the finger. Since much of the work is not unimodal, where 46 out of the 119 FADs sporting any kind of input are in fact multimodal, the counting in the

(a) Finger flexion sensor [67] (b) iRing [158] (c) uTrack [33]

Figure 2-6: Input modalities in finger-worn devices. (a) a sensor for the flexion of finger tendons, (b) an IR reflectance sensor for different parts of the finger's skin, (c) a magnetic sensor. Images courtesy of authors.

following list includes duplicates. To make a further distinction, we do not classify the underlying sensing technology itself but rather the outcome input signal that creators used to support the intended interaction.

The following input modalities were recorded in our survey:

- binary-state buttons [44 instances]
- inertial: translation (accelerometers) and rotation (gyroscopes) [39 instances]
- imaging: cameras or other photometric detectors [28 instances]
- pressure or force, against a surface, other fingers or objects [22 instances]
- proximity, to other devices or limbs [15 instances]
- 2D positioning: joysticks or touch-sensitive pads [13 instances]
- acoustic: microphones or general sensing of the audible range [10 instances]
- biometric: pulse, oximetry, blood pressure, etc. [9 instances]
- rotation, of the FAD device against the finger [5 instances]
- magnetic: Hall effect sensors or other magnetometers [4 instances]
- thermal [4 instances]
- bending [3 instances]
- mechanical, coded gears [3 instances]

Binary buttons are widely used in FADs for their straightforward operation and user familiarity, where the buttons are usually located on one finger (facing out or in) and operated with the thumb. As is the case in most early FAD applications the major usage was to create finger-wearable keyboards [189; 116; 113; 85; 12] and mice [245; 80; 94; 14; 231; 199; 191]. Other prominent usage profiles for binary button include communication [48; 131; 218; 145] and assistive technology [46; 66; 184; 153].

Input: Buttons

Pros: Binary state buttons are *inexpensive*, easy to *integrate* and simple to *operate* for the user. They are relatively small, so can be placed almost anywhere on the finger.

Cons: Buttons provide only a *simplistic binary operation* with little information for the system. They are also not made as thin as touch-sensitive surfaces (e.g. capacitive).

Main applications: computer input (keyboards, mice) and communication.

Inertial measurements units (IMUs) are implemented using accelerometers (to sense motion) and gyroscopes (to sense orientation). Much attention was given to creating keyboards by detecting finger taps [48; 49; 171; 108; 100; 88; 105; 125; 182] and mouse-like input [239; 163; 264; 71; 176; 121; 168; 151; 236; 156] for the clear affordances these sensors provide in detecting abrupt motion or integrating it to a velocity signal. However recently, gestural interaction has become the de facto usage scenario of inertial measurements in FADs with an abundance of recent work [112; 84; 96; 83; 229; 252; 125]. However additional usage scenarios do exist, especially in the biometrics domain where the acceleration signal is traditionally used to filter noise from arterial blood flow measurements [8; 201; 60; 115; 74].

Input: Inertial (accelerometers & gyroscopes)

Pros: IMUs are relatively *inexpensive*, *small*, and work very *intuitively* with finger or hand air gestures.

Cons: Sometimes IMUs require calibration, and cannot reliably measure precise translation. Gesture recognition requires *additional computation power*.

Main applications: cursor control, gesture recognition.

Finger worn cameras (or other photo-reactive elements) combine a powerful sensor with a highly sensitive body part (see Fig. 2-7). For this reason many researchers use them for assistive technology work, looking to recapture even a sliver of a lost sense of sight in a variety of ways: reading text or detecting patterns in print [6; 66; 111; 203; 226], detecting objects and scenes [184; 153], navigation and general sightless usage [72]. Nevertheless, finger worn imaging was shown to be useful for screen cursor control [263; 257; 236; 200; 168; 97; 174], natural interaction with objects [141; 257; 184; 175; 68], as a wearable barcode scanner [232] and simply as a wearable camera [148]. In all recorded cases the imaging sensors are positioned facing away from the finger and pointing forward in the direction of pointing or down in the direction of touch.

Input: Cameras and photodetectors

Pros: Cameras provide *high-dimensional input* and enable object or scene recognition, complex hand gesture poses. Photodetectors are *small and cheap.* Can work beyond the visible spectrum, e.g. in the dark.

Cons: Analyzing camera images or video requires *considerable computation* and *power*, they are also usually quite *large.* Photodetectors provide a *low-dimensional*, albeit continuous, signal. Finger-based imaging is *not intuitive* to the user.

Main applications: natural interaction, assistive technology.

(a) EyeRing [153]

(b) FingerReader [203]

(c) FingerSight^{TM} [72]

Figure 2-7: Finger-worn cameras. Images courtesy of authors.

Pressure sensors were used to create virtual reality interfaces to bridge the tactile sensation gap of the virtual world by detecting the force applied to fingertips [133; 170; 34], or to restore a lost tactile sensation in the fingers (in people suffering from Multiple Sclerosis) [82]. In the department of computer interaction, Zloof et al. postulated a rotating ring to control cursors [267], Xiong et al. used pressure sensors to create a thumb-mouse [253] and Chatterjee et al. a finger-worn keyboard/mouse based on 8 pressure sensors [32]. While most finger-worn pressure sensors are based on layered electrodes, Ogata et al. have created a force sensitive ring based on infrared lighting and photodiodes, taking into account different reflectance properties of human finger skin [158].

Input: Pressure sensors

Pros: *Intuitive to the user*, for the natural heightened sense of touch in the fingertips. Usually provide a continuous signal with more information, at low power consumption.

Cons: Only support proximal activity, i.e. cannot detect gestures.

Main applications: virtual reality, computer input.

Sensing modalities such as 2D sensors (miniature joysticks, touchpads, pressure pads or ball rollers) were mostly used to create keyboards and mice [144; 191; 119; 15; 44;

253; 95; 94; 14]. Sensing the bending of fingers, thermal sensors and pulse sensors showed potential in biometric applications [120; 67; 8; 201; 181; 60; 249; 115; 74], while mechanical fixtures attached to the hand were used to create virtual reality interfaces with sensing as well as feedback [56; 161; 223]. Much of the other modalities (microphones, proximity sensors) are combined with other modalities, however Chen et al. created a finger-wearable 3D input device[33] and Maekawa et al. a hand held devices sensor [128] using only a single magnetic sensor.

2.2.3.3 Embedded Output Elements

Output in a FAD is mostly geared towards the user, providing notification on a performed manual action or output from an external system, delivered to the finger for its sensitivity to many kinds of energy and bodily visibility. We consider two classes of output FADs, ones with Human-Detectable Output - HDO (i.e. energy detectable by the innate human senses), and Non-Detectable Output - NDO (e.g. magnetic energy, or radio). Even though HDO FADs were extensively explored, FADs in general do not have HDO, where only roughly half of the works we surveyed had any kind of output modality detectable by human senses. An interesting special case are the pure-HDOs (i.e. FADs without any input modality) that are prominent in creating interfaces for virtual reality, where the focus is on stimulating the fingers to feel virtual objects rather than sensing the world [114; 195; 197; 143; 99; 103; 196; 54]. Other prominent outlets for HDOs are naturally communication, in the form of a mobile phone companion device [145], and assistive technology, in the form of wearable finger-braille devices [70]. So far, we only encountered a single finger worn device concept based on temperature-output [4].

NDO FADs are mostly used as means of input to other devices, mostly for cursor manipulation, by utilizing a companion module such as a wristwatch [64], a bracelet [35] or external light sensors [214], to pick up on the undetectable emissions (or reflections) from the FAD.

The following are the dominant output modalities we encountered in the surveyed works:

—	HDO	Vibration	[30 instances]
—	HDO/NDO	Light	[22 instances]
—	HDO	Tactile (other than vibration, e.g. compression)	[19 instances]
—	HDO	Display (complex lighting setup)	[11 instances]
—	HDO	Audio	[5 instances]
—	NDO	Radio	[4 instances]
—	NDO	Magnet	[3 instances]

The primary HDO modality is vibration, with 25 unique publications reporting the usage of vibration capabilities in a FAD (due to the evident widespread use of vibration in FADs we separate it from other forms of tactile feedback). Presumably, this is owing to the fact that human fingertips are highly sensitive to vibration through mechanoreceptors embedded close to the skin and deeper inside the finger, allowing to effectively detect vibration in the 10-300Hz range with miniscule displacement (10 microns or less); other parts of the finger are slightly less sensitive than the fingertips as the concentration of mechanoreceptors decreases [233]. The two major themes in using vibration output is for assistive and communication applications. Usage of vibration in assistive applications focused on finger-braille [70; 5; 134], sensory substitution [6; 203], sensory amplification / mediation [82; 106], and in communication we find a large number of smartphone companion devices that alert of incoming messages and allow for rudimentary response [131] (in interest of brevity, we invite the reader to view the complete list in the appendix). Other applications for vibration include gestural and cursor control [200; 125] and virtual reality interfaces [114; 161].

Output: Vibration

Pros: Vibration motors are relatively *cheap* and easy to *integrate*. Fingers *easily detect vibration* at multiple degrees of frequency and amplitude.

Cons: The entire finger is not as sensitive to vibration as the fingertips. Motors can be *large* and may draw significant *power*.

Main applications: communication, assistive technology, virtual reality.

In contrast to vibration, static and near-static (<10Hz) tactile feeling on the skin is detected with a different set of mechanoreceptors that are distributed farther apart on the surface of the glabrous skin (that of the hand) [Johansson 79]. Non-vibratory tactile feedback via FADs is virtually dominated by applications for virtual reality, for the goal in such interfaces is to simulate the force feedback from grasping or touching virtual objects. While a great number of actuators for tactile feedback exist (electromagnetic, piezoelectric, electrostatic and others [16]), their integration in a finger wearable form is not trivial. To apply compression and shear force on the fingertips, researchers explored using miniaturized finger-mounted DC motors and strings [170; 196; 34], wrist-mounted motors with companion finger-worn mechanical pads [223], motor-driven extruding nuts [93] or shape-memory alloys [105].

Output: Tactile

Pros: Fingertips are very sensitive to tactile response even at very subtle levels, with good resolution and separation.

Cons: Mechanical constructs to provide pressure or compression are large and power hungry, although alternatives exist (e.g. piezoelectric, electromagnetic).

Main applications: virtual reality.

Simple finger-worn light output is a trivial augmentation of the finger already available in products targeting, for example, recreational usage [147] and the aviation industry

[109]. However in FADs, which require a non-trivial augmentation, light is usually intended to deliver information visible to the wearer's eye (HDO) or with sensors accompanying the light source on the FAD (NDO). Many FADs use a single light source in communication applications to indicate pending operations or messages [144; 107; 183; 169]. Other uses for light output are reporting of operational status such as charge, power on/off or internal state [239; 84; 200; 125; 151], to visualize gesture [96], or a laser to indicate pointing direction [263]. NDO light was used for example in the iRing [158] where a NIR light source was used to detect different skin regions.

Output: Light

Pros: Simple to integrate, intuitive to understand, cheap, has a small footprint and low power consumption. Can easily work as a non-detected output (NDO) outside the visible light spectrum.

Cons: Cannot provide a large amount of information.

Main applications: communication, lighting.

Displays as opposed to discrete lighting sources can deliver a much higher order of HDO information, despite that, the domain of wearable displays is still in its infancy and FADs packing displays are at the forefront. Usage of finger-worn displays mostly revolve around communication and organization, where they display caller ID, a message or calendar information [78; 218; 145]. However, interesting examples also include a nail-worn display to overcome the finger's occlusion of touch-screen devices [229] (see Figure 2-5a), and a palm-mounted screen to assist in hands-free operation for example while driving [113]. Additional HDO modalities in FADs also include audio, where speakers create a finger-based telephone handset [50; 47] or as an additional feedback modality [84].

Output: Display	
Pros: Can deliver a large amount of information, and utilize "dead space" for augmentation or occlusion.	**Cons**: Large and power hungry, and often are low-resolution at small sizes. Require more computational power to drive. Applications are limited without touchscreen capabilities.
Main applications: communication	

While finger-worn NDO is far surpassed by HDO in frequency of usage, some did use magnets in coordination with other wearable sensors to control cursors [64], create 1D input [9], or control personal devices [240]. Vega and Fuchs also postulate finger-wearable RFID to control and communicate with devices through the finger [240], while others proposed to use finger-wearable RF antennas [208; 244].

2.2.3.4 Where the Action Is

One of the most complicated classifications of FADs is based on their intended action and where it takes place. Based on the collection of works surveyed, we determined the following classification for the action: (see figure 2-8)

- Clicking or touching the device itself [58 instances]
- Touching a surface with the finger [33 instances]
- Pointing or external action [29 instances]
- Gesturing the device in the air [25 instances]

However some of the instances offer more than a single action to make a combination, and the most common combination being Pointing + Clicking, for example in

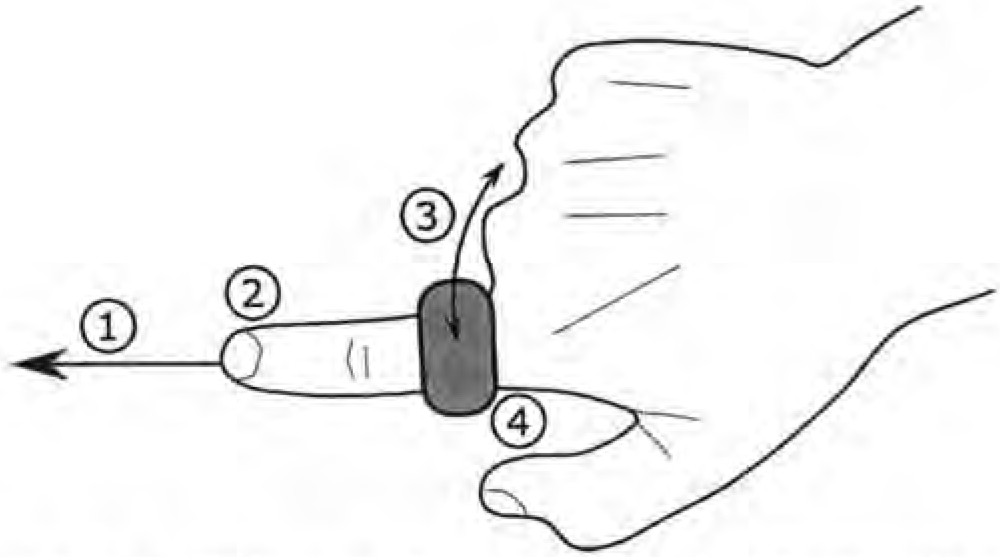


Figure 2-8: Types of FAD actions: (1) Pointing / External, (2) Touching / On surface, (3) Gesturing, (4) Clicking / Touching on device.

the finger-wearable remote control from [239]. Pure output FADs without an input modality cannot fit into this model and therefore are not classified in any category.

Pointing is a cardinal deictic gesture in our gestural language, it is cross-cultural, usage of it dates as far back as ancient cultures worldwide, and is exhibited even in infancy [139]. It is therefore a very convenient platform for augmentation, and was detected as such by many creators of FADs. The pointing gesture usually suggests the existence of a referent in the immediate environment, thus several pointing-direction recovery systems were suggested: global localization in the environment [112], local based on instrumented beacons or sensors (Infra-red: [141], Ultrasonic: [201], Magnetometers: [64; 121; 29]), local based on fiducial markers or natural features [257; 175] or integrative based on accelerometers [71]. Other pointing augmentations do not make use of a localization mechanism when the interaction is oblivious to the spatial domain, simply examining the referent ahead [263; 184; 68; 153; 194; 72].

Action: Pointing

Pros: Socially acceptable, cross-cultural and natural gesture, which also provides ample information for the system, as well as others, to understand the user's intent. Can leverage the finger's flex, or bend, for an additional signal.

Cons: The sensors must be aligned carefully with the finger, which presents problems of mounting, calibration, occlusion (by the nail or fingertip) and accommodation to warped fingers. May constrain the FAD to be placed on the index finger. Some pointing gestures carry negative social meanings.

Main applications: appliance control, assistive technology.

The most common action modality is clicking or touching the FAD, and is most commonly done by adding a button or other interaction elements to the device body. The opposing thumb easily reaches a button located on the side or bottom of the FAD (depending on the wearing finger) to support a single handed usage, which was the goal in many recent products [156; 151; 145] and in academia [185; 68; 153]. More subtle input, or one that requires a gradient of values, was done via pressure sensing [267; 44; 253; 32; 82; 158; 182] or a touch pad on the device [119; 144; 253; 242; 15]. Zloof et al. suggested using the rotation of the FAD around the finger as an on-device input method as far back as 1996, and years later it was picked up by [9; 158].

Action: Touching the device

Pros: Easily understandable for the users, mostly easy to implement and robust. May allow for discreet, unseen function, even with a single hand. Can be placed virtually anywhere on any finger.

Cons: Many times uses a binary signal, i.e. a button. Oblivious to the environment, narrowing the interaction to the device itself.

Main applications: communication, computer input.

Touching the surface when wearing a FAD was markedly used for cursor manipulation or a chording keyboard since the early days of finger augmentation [48; 171], and this still trends with very recent work such as [257; 97]. Leveraging the fingertip's very sensitive tactile sensation was targeted by creators of assistive FADs, guiding the process of scanning with the finger [6; 203; 226], for sensory substitution [194] or using the body as the input means [146]. Prattichizzo et al. explored the sense of touch in the context of a mixed reality system [170; 34] using tactile actuators alongside pressure sensors that gauge the real vs. generated force of touch. Other have explored touch as a gestural interface to enhance the interface of everyday objects [40; 252] or displays [121].

Action: Touching the surface

Pros: Functions in conjunction with the sense of touch, and may also provide the system information about the touched surface or the "tapping" gesture. Lends itself easily for mixed-reality applications.

Cons: In some implementations can obstruct the inherent sense of touch, and may warrant specific positioning on the finger due to the high sensitivity of the index fingertip.

Main applications: assistive technology, virtual reality.

Using the FAD to detect gestures is a recent addition to the FAD interaction milieu, where the postulation to augment fingers to use gestures as an input was set forth in close vicinity by [222; 100] and [239]. Although the usage of gesture sensing technology existed in FADs since the 1990s [48; 171], it was not used to detect gestural motion rather finger taps for keyboard input [108]. In 2006, SourceAudio already introduced a product featuring a finger-wearable wireless gestural interface for adding effects to an electric guitar [225]. Following was a wave of interest in gestural interaction both in academia and in the form of patents, which culminated in a number of products released in the last 2 years [236; 125; 156; 151; 182]. The academic front explored usage of finger gestures for appliance control [112; 84], as an input device to replace a mouse or keyboard [88; 64; 163; 71; 33; 83], or detecting gestures of a novel vocabulary [96].

Action: Gesture

Pros: Leverages on the dexterity of the fingers and the recent wave of interest in gestural interaction, which makes it fairly understandable to the user. Implementation is cheap, power efficient and can be made wireless.

Cons: Often constrained to a set of canned gestures, and otherwise may require calibration or training. Some gestures may create awkward social situations, especially as they are clearly visible to others but in fact made in private, and carry little conversational meaning.

Main applications: computer input, appliance control.

2.2.4 FAD Applications

As we discussed in the last sections, FAD instances present plenty of input and output modalities with a wide span of capabilities. These instances cluster together in a number of application domains that look to achieve a common goal–enhancing or enabling a manual operation, gesture or inherent ability. Applications for FADs trend with the eras in correspondence to the larger trends in the world of HCI. The current trend is quite clear: moving from keyboard, mouse and/or remote control input to gestural and natural interaction. Similar trends in application were noted in the past, for example with FADs as virtual reality interfaces, which have seen a wave of interest in the beginning of the 2000's and scaled back a few years later. This section describes the major and minor application categories that materialized from the body of work.

The following are the major application domains we recognize for FADs:

- Mouse-like cursor input [42 instances]
- Communication [22 instances]
- Assistive technology [21 instances]
- Appliance control [21 instances]
- Gestural interface: in-air gestures, hand pose, pointing, etc. [21 instances]
- Virtual reality interface [18 instances]
- Keyboard input and output [16 instances]
- Biometrics [11 instances]
- General: timewatch, jewelry, camera, etc. [11 instances]
- Natural interaction with objects [10 instances]
- Social interaction [5 instances]
- Industry [4 instances]
- Learning [4 instances]
- Creative interaction [4 instances]

2.2.4.1 FADs as paired input devices

Most FADs are input devices that work in tandem with existing electronic devices: personal computers (e.g. desktops or laptops,) mobile devices (e.g. smartphones, tablets or wearables,) and even home appliances (e.g. TVs, sound systems or kitchen appliances). The goal of these FADs is to offer more efficient or natural interaction with paired devices through an always-on, steerable and light finger-worn device. It appears FADs offer interaction around several recurring themes: keyboard input or mouse-like cursor input [236; 191; 174], remote or gestural control [84; 156; 182; 125] and communication [47; 131]. Beyond interaction with traditional personal devices, FADs are customarily used as a paired output device for virtual and mixed reality systems offering stimulation in the hand [195; 161; 170; 196; 54]. In the last few years however, we observe a rising prominence of the notion of natural interaction, which we define as an interface with non-instrumented everyday objects (for example [128] who are using a magnetometer). Letting FADs leverage on the well-practiced pointing gesture to steer the interaction [153], the FADs can be a cursor for the main device for further action or computation.

2.2.4.2 Assistive applications

The assistive technologies domain presents its own set of challenges for FADs, arising from the special needs of the target audience [241]. This may also be the reason concept designers are drawn to this type of user interfaces, to present a wishful and critical outlook on the possible role of FADs as assistive technology. Hedberg and Bennett envision a whole finger augmentor with multiple sensors and capabilities: Braille tactile screen, camera for text-recognition, buttons, microphone and wireless connectivity [66]. Lee brings a more modest vision of a fingertip augmentor that reads barcodes and wirelessly delivers useful information to an earpiece [111]. In the engineering world, some work was dedicated to creating a wearable form of finger-braille: an adaptation of the braille to 3 or 5 fingers coded-sensations instead of the usual printed raised-dots [70; 5], while others focused on accessing visual information, in particular printed text [203; 153; 226]. More attention was given to tactile displays

[104; 103] and enhancing the tactile sensation in the fingers in cases where this sense was impaired, for example helping persons with multiple sclerosis [82], or helping people whose work demands high dexterity (such as surgeons or assembly) [106].

2.2.4.3 Biometric applications

In the biometrics domain, the most prominent examples of finger augmentation are the ubiquitous pulse oximeters that are donned by hospital patients to monitor their vitals. These have been in existence for many decades, and recently have reached wide usage in medical facilities with a range of available products [157]. Their central mode of sensing is photometric, and relies on the different light reflection and absorption properties of oxygen saturated and unsaturated hemoglobin within the bloodstream. As the field of finger-worn pulse oximeters seems to have moved out of the academic world we will not review its history but rather new explorations of this type of devices. Some of the challenges recent implementation try to cope with are: sensing pulse while in motion [115; 74], wireless and low-power operation for long-term sensing [181; 60] or additional sensing modalities such as temperature [8; 201], proximity [166] or posture of the hand and fingers [73].

2.2.4.4 Industrial applications

The industrial domain seems to remain somewhat indifferent to the outburst of finger-worn devices except for solving very specific needs in manufacturing and operations. Nevertheless, we could find: a wearable controller for a sewing machine [192], an industrial-grade finger worn barcode scanner [232], a ring to prevent the misfire of firearms [17], a human-robot interaction and guidance device [132] and a device that enhances the tactile sense in finger-inhospitable environments such as cold and dampness [106]. Notwithstanding, we can clearly see how most of the FADs in other application domains potentially have impact on the industrial world with trivial adaptation.

2.2.4.5 General usage and fringe applications

Looking away from the major augmentation theme, we find an interesting set of applications for various types of augmentation: a finger-watch [46; 31; 140; 218], a finger-camera [148; 51] and jewelry [183; 144]. We also recorded an interest in the social aspect of using a FAD, with researchers creating a discreet interaction application [9; 97], and applications to stay connected with loved ones [112; 23; 249]. Another interesting minor line of applications for FADs are learning and creativity, for example in playing the piano [75; 102] or the guitar [225].

2.2.4.6 Challenges and Opportunities in Designing useful FADs

FADs are relative newcomers to the world of wearable computing, where much was already tried and tested. Nevertheless their unique traits help designers approach old problems with a new toolkit. In table 2.1 we list a number of challenging problems in interaction for which FADs offer unique support. The rest of this section is devoted to design considerations one could follow when thinking up new FADs.

2.2.5 Design Considerations

Designing useful FADs depends greatly on how the creators incorporate the finger's senses and function into the interaction. We compiled a list of considerations to contemplate when designing a new FAD, that also exposes new opportunities to seek out under-explored territories. With this list we also stress the wholesome approach one should take when designing for the finger, for the many and intricate aspects of

	Challenge	Opportunity
Immediate	An interface the user can operate at any time with minimal effort.	FADs are worn on the body's most dextrous limbs - the fingers, and bring the point-of-interaction literally to the user's fingertips. They can reduce the reliance on external setup, and sense at a very high resolution.
Close-up	Sensing at close proximity to the user, utilizing fine motoric skills.	FADs can be placed on the index finger for minuscule yet very precise motion. Utilizing the fingertips for more than close-up tactile sensing is under-explored.
Discreet	An interface that is private to the user, can be made unobtrusive and inconspicuous.	Placing I/O elements on the inside of the FAD, facing the palm where the thumb can easily operate, creates a private interaction space. Using small and weak tactile response actuators can output information strictly to the wearer and maintain the natural function of the finger. Ring FADs can be perceived as socially acceptable and not raise attention.
Subtle, Efficient	An interface that outputs in low-magnitude, low-power and high-resolution.	The fingers are highly sensitive in frequency, magnitude and phase (separation or translation of stimuli), especially around the fingertips, allowing to use a weak and efficient form of output (e.g. tactile). The hands are also very visible to the user, allowing for visual feedback.
Assistive, Augmenting	An interface that assists in the case of impaired senses, limiting situations, or to augment the inherent human capabilities.	The fingers are already used as substitute eyes, ears and mouths, which makes FADs a prime candidate for assistive applications. FADs can translate from one modality to another in high fidelity, bringing the input and output together on one body part.
Gestural	An interface that works by natural gestures.	Starting at a very young age, hands and fingers serve as one of the central means of gestural language. This makes gesturing with a FAD easily understandable to the wearer, and less awkward.
Bio-sensing	An interface that monitors biological signals from the body.	The fingers have a dense network of nerves and blood vessels, allowing to externally inspect some aspects of the bloodstream (e.g. photoplethysmoraphy), the sympathetic nervous system (e.g. galvanic skin response) and others.
Multipurpose, Repeatable	An interface with a broad range of utility, application and reusability.	Fingers are naturally used for a wide range of both day-to-day and special activities, thus FADs may be used to sense and augment them using the same form factor.

Table 2.1: Problems in interaction and the opportunities FADs offer to scaffold them.

the finger as a living limb and an object of meaning. The list emerged both from the surveyed body of work and our own explorations in creating FADs. One should note the most trivial of considerations in creating a FAD – making sure it achieves the intended operation – is not discussed in this list, we chose to rather focus on the more latent and underserved design aspects. These aspects have nevertheless resonated with many FAD creators, and we regularly point to the relevant works, however we wish to put them in a single instantiation that can serve as a guideline.

2.2.5.1 Using the anatomy of the finger

Fingers are incredibly sensitive to a number of types of energy, they are the most dexterous and strategically positioned limb, and they are highly represented in the primary somatosensory and motor cortices in our brain. These and other anatomical properties, widely researched outside the HCI community, turn fingers to be a boon of interaction and augmentation potential. Researchers of novel interaction methods discovered some anatomical traits in the exploration of: the finger and fingertip highly dense mechanoreceptors [82; 106], the flexion of finger tendons [67], bone conduction [47] and compression properties of the soft tissue [133], however the vast majority plainly use the tactile sense via vibration. The finger is also highly sensitive to changes in temperature via thermoreceptors, however only [44; 8; 201; 96] have discussed it as an input modality and [4] for output. Established usage of the finger's anatomy in FADs was recorded for pose detection (via a mechanical [56] or other sensing modality [73]) and for photoplethysmography [115], however an interesting yet somewhat underexplored territory is that of proprioception or sightless action [9; 160]. Large areas of design for the finger's anatomy remain at large for exploration: Thermal, Nociception (pain), Irritation, Perspiration and Humidity, and more. This evident gap poses an immense opportunity for designers to deepen the understanding of the finger's physiology for usage in augmentation.

2.2.5.2 Using well practiced behavior

The importance of the finger as a primary tool for sensing and interacting with the world is uncontested and heavily relied on by UI designers. However often designers neglect the fact that fingers play a central role in our gestural and behavioral language from the very moment we are born, and such practiced behaviors carry deep meaning [139]. Deliberately leveraging practiced behavior played a minor role in finger augmentation so far, with works augmenting the pointing gesture [239; 141; 112; 153], other common gestures (the "phone" [47], stroking [114], scratching [166]), daily activities [83] or holding an object [252]. Still different kinds of gestures and behaviors remain to be explored in the context of FADs, for example iconic gestures (representing an operation, such as "cutting" or "chopping") and metaphorical gestures (such as a "speaking mouth"). In the work we surveyed there is evidence that finger augmentation could benefit from using practiced behaviors of the hand and fingers to invite the users into a recognized interaction with the world, rather than introduce a new manual operation. On the other hand, augmenting the fingers could impede a routine operation, such as washing hands, handling a manual tool or playing a music instrument.

2.2.5.3 Using the ring as a fashionable traditional object

Finger rings are objects of tremendous tradition, as jewelry and symbols of stature, power and bond. Their history is believed to run back to the beginning of mankind, however the concrete evidence of finger-worn fashion dates to only a number of millennia ago. Rings of significance appear throughout the narratives of ancient cultures (Egyptian, Greek, Roman, Israelite, Persian and Chinese), and are aptly represented even in the narratives of today [137]. This rich backdrop to our interest in finger augmentation is lightly touched upon in the realm of engineering, however it is starting to take a more prominent stance with the rise of commonly used wearable computers. A new project named Ringly [183] is specifically designed as a smart finger jewelry, however concept designers of finger augmentors already discussed the aspect of finger-worn fashion in the past [144; 100; 31; 240]. Beyond fashion, rings hold a variety of symbolic meanings, such as engagement or belonging to a group. Building on such symbolic meaning is practically non-existent (although not unheard of [249]) in the field of finger augmentation, and presents one of the most exciting areas for investigation.

2.2.5.4 Creating a comfortable usable design

Comfort and an appealing form factor are a cornerstone for successful design for the body. In augmenting fingers with devices this is of highest importance, since fingers and hands are very sensitive and a very visible body part. Naturally, no two fingers are the same size (girth, length) or shape, however generalizations for these aspect were proposed in the form of finger size charts. Less standardized are the wearing and removing mechanisms, which are equally important and disregarded by the majority of FAD designers. Wearing and removing mechanisms come in a range of types: simple rings [145], clasps [191], unclosed rings [182], flexible or rubber fastening [203] and others, however unfortunately it seems the prolific way of mounting components to the finger is to do so without care. Placing components should also work to the function of the device, for example buttons for the thumb should be placed on the side [153] and light output will be most successful on a line of sight to the eye, i.e. on top [96].

2.2.5.5 Using a companion device, or the FAD as a companion device

Often FADs are not the only device the user would wear to perform the intended action, rather it works in tandem with an external device, or the FAD itself is made from more than a single finger-worn device. Many cases display rings that are wired to a device on the palm [100], wrist [49; 264], or more commonly a connection to an external non-wearable device [97; 203]. Wireless FAD are found in abundance, however this does not mean they are used solo, in some cases two FADs are required: on the thumb and another finger [36; 33] or one on each hand [71]. A wireless single FAD would be the least cumbersome, but it poses both a technical and an interaction challenge to fit the components in a tight space and achieve an action requiring more than a single augmented finger. Using a companion device, e.g. a smartphone, may circumvent these obstacles.

2.2.5.6 Assistive augmenting technology

While a formidable amount of work was devoted to creating assistive finger-worn devices, it is not the mainstream agenda in finger augmentation, which creates ample opportunity. One important aspect to notice is that fingers are traditionally more than just fingers for people with different impairments, they often are substitute eyes, ears and mouths. This introduces a dual challenge: not obstructing the inherent function of the finger as a substitute sense, and adding a meaningful assistive augmentation. These considerations were discussed shortly in the context of finger augmentation [153] and naturally much deeper outside of it [241], with the major lesson being that assistive devices should be useful and unencumbering to become successfully adopted. The majority of the work surveyed was geared towards assisting people with a visual impairment [203; 226] and only little to other impaired senses or conditions [166; 82]. To expand the reach of finger assistive augmentation, it is useful to observe the wider range of assistance, as one particular technology could also be practical outside its intended domain of application and target audience.

2.2.6 Conclusions

2.2.6.1 The rise of FADs

Finger augmentation is on the rise as a domain of user interface, as well as a new branch of wearable computers that taps into new types of sensing and signaling. Recently FADs attracted considerable interest in academia but also as a commodity through products and projects, with new work contributed to the pool yearly. This success can be attributed the re-discovery of fingers as a comfortable space for augmentation via electronics, building on past traditions. Fingers as a driver of focus, the sense of touch, and both deictic and iconic gestures, offer easy access for user interface developers to the body language and manual actions of the wearer. With wireless sensor technology becoming ever smaller, efficient and simpler to integrate, the small form factor of a finger wearable is no longer a deterrent to creators.

The tradition of finger wearables reinforces the belief that FADs could be mainstream devices of interaction and have a mass-market feasibility. The prominent recent application domains of FADs are user input to a computer system (personal computer, mobile device or smart environment), suggesting the future lies in discreet, fashionable immediate-control devices that target end-users as their audience. Sleek design and omni-connectivity articulate the timeless narrative of jewelry doubling as objects of power or function, which is a central theme in contemporary user interfaces.

2.2.6.2 Avenues for future research into FADs

While it is hard to predict the future of FADs, the current trend is showing a promising outlook. Wearable computers, now becoming a commodity, contribute to public interest in finger worn devices especially around complementing personal mobile computers such as smartphones. While products are already going to market [218], this enterprise is far from complete since the technology (input, output, power, and connectivity) for the ring form factor is immature and can benefit from further research.

Assistive technology with finger wearable devices is still in its infancy, however the buds of progress are prominent [205; 226]. This domain also presents a wide range of opportunity to prototype and research (see Table 2.1), particularly in the areas of sensory substitution, enhancement and recovery.

Also in evident pressing need, is deeper research into leveraging the anatomy and the natural behavior of the fingers, as currently these considerations are somewhat overlooked. This may be another effect of a vertical research agenda that gives less attention to cross-pollination between disciplines. Knowledge of finger anatomy is extraordinary rich in the traditional disciplines of medicine and physiology, however it is fairly unreachable from the human-computer interaction perspective. An integration of this knowledge, in form of demonstrated guidelines and prototypes, will certainly be a gift to both courses of research.

2.2.6.3 Contributions and summary

Over the past years we carried out a comprehensive survey and overview of the work on FADs. We created a categorization framework that distinguishes FADs based on their key elements: form factor, input modality, output modality, interaction and application domain, where each element was further scrutinized to sub categories.

Our methodological review also resulted in a list of challenges FADs designer face and opportunities to overcome those. Nevertheless, we believe further inquiry into understanding the holistic nature of FADs as personal devices is required, attributed to the rich history of finger-worn objects of meaning and fashion. Research into the symbolism of rings and fingers is paramount in wearable artifacts design (jewelry for instance), but it did not yet percolate into engineering of augmenting devices. To achieve this integration, more guidelines that bring ergonomics and fashionable design factors into technological prototyping must emerge.

2.2.6.4 Continuation of the document

The next chapters describe our own contribution to the world of finger augmentation in form of design probes, research projects and evaluations. We begin in Chapter 3 with our first exploration of augmenting the natural pointing gesture in the EyeRing project using a finger-worn camera. The following chapters elaborate on the idea of a seamless connection the augmented-finger creates with the environment, offering additional functions: reading text (with the FingerReader, Chapters 4 and 5) and musical notation (with the MusicReader, Chapter 6). Throughout the document we revisit some of our suggested design concepts for augmentation presented in this chapter: building on familiar gestures, using the natural senses of touch and proprioception, and a cross-domain inclusive design.

Chapter 3

The EyeRing

A Finger-Worn Visual Assistant

This chapter presents the design and implementation of a finger-worn I/O device, the EyeRing, which leverages the universal and natural gesture of pointing. Through demonstrated use cases of the EyeRing for both visually impaired and sighted people, we suggest that finger-pointing augmentation may indeed offer a more seamless interface to the immediate surroundings than using a standard mobile device. We also report on a user study that demonstrates how EyeRing reduces effort and disruption to a sighted user. We conclude that this form factor and interaction modality

Figure 3-1: EyeRing: A finger-worn input device. (a) EyeRing prototype. (b) CurrencyDetector application. (c) TagDetector application. (d) Interaction with printed media.

offers enhanced, seamless interaction with objects in the environment by relying on practiced human behavior using a small wireless device.

3.1 Introduction

The pointing gesture is fundamental to human behavior [136] and used consistently across cultures [138]. It begins at an early developmental stage [27] and lets humans reference proximal objects as well as abstract concepts in the world. Pointing gestures, which are an inseperable part of our gestural language, are inherently used for *interaction.* This is a strong motivation for designing pointing-augmenting devices for interaction: not only can we build upon a natural and universal human behavior, but we also benefit from the focus and intentionality in the gesture itself. Given the recent increased interest in using speech to interact with mobile devices [216], it is a logical next step to support a user in pointing at an object while stating a comment or asking a question about it.

Visually impaired persons may also benefit from using pointing-gesture interfaces, although the motivation for this stems from a different perspective. Turning to recent literature on interface design for the visually impaired, we note three desired qualities: assistive technology should be *socially acceptable*, work *coherently for disabled and non-disabled* alike, and also support *independent and portable interaction* [209; 210; 250; 173]. The finger-worn device presented here follows this design paradigm: it looks and offers the same affordances and mode-of-use to both sighted and blind users, as well as used in a mobile, self-sufficient way. We deepen the discussion of design guidelines for sighted and visually impaired in the 'Design Considerations' section.

Among their many meanings, pointing gestures are perhaps most regularly used for referring to a place or a thing in space. We propose a method for augmenting the pointing gesture for information retrieval tasks. Previous research work in the field of augmenting pointing gestures revolved around control [238] and information retrieval [142]. These works and others utilize a specialized sensor between the pointing finger and the target, such as an infrared connection in the work of Merrill et al. [142]. This implies a pre-rigged environment, which inhibits natural interaction and limits the domain of use. We therefore chose to use a visible-light spectrum camera as the sensor. Naturally, this means the computational element of the system is more complex, as we use computer vision methods to extract information from images.

The EyeRing is an input device consisting of a camera mounted on a ring (worn on the index finger to follow the natural pointing gesture) with an embedded processor and wireless connection to a computation device (typically a smartphone). We employ a number of computer vision techniques to recognize objects or locations based on one or more images taken by the ring. The EyeRing system also has a speech recognition service to enable voice commands, and speech output (as well as screen output) as the means of communicating information. This platform enables a whole series of applications in which a user may acquire information or control objects/spaces in their proximity. For example a blind user could simply say 'currency' and point at a currency note to hear the value of it.

3.2 Related Work

For a comprehensive overview of finger-worn augmentation devices the reader is referred to Chapter 2, however we hereby briefly mention a number of works that preceded the EyeRing and inspired our work.

Leveraging the pointing gesture with the application of remotely controlling objects in the environment was implemented in the Ubi-Finger [238] and FieldMouse [215] projects. Efforts to attach and retrieve information from physical objects by pointing were implemented in [142] and recently in [256] using IR beacons and coded textures. However these applications often require the environment to be instrumented with sensors and markers, which limits the interactions to instrumented environments. Horie et al. overcome this by using accelerometers to augment the two index fingers simultaneously and triangulating the intersection of the respecting finger-pointing rays [71].

In the assistive technology domain, FingerSight [53; 263] provides a visual-tactile substitution system by converting visual information in the direction of pointing into feedback, which is also embodied in the type of interaction suggested by EyeRing.

3.3 Design Considerations

The design of Eyering follows guidelines set forth by Rekimoto [179] as well as by designers of assistive technology [209; 250].

3.3.1 Designing Interaction With Finger-Worn Devices

Rekimoto's work on *Augmented Interaction* proposed a number of guidelines for designing unobstructive wearable technology: *straightforward operation, using real-world situations as implicit input*, and ensuring that the technology is *socially acceptable.* These guidelines suggest that users should be able to operate the device without holding it and the device should allow for "quick changes between normal and operation modes." The goal of "[using] real-world situations as implicit commands" is embodied in our work through the use of a camera and computer vision methods. In addition, the device "should be as natural and (conceptually) unnoticeable as possible for use in various social settings." EyeRing is designed to look like a common wearable accessory, a wireless ring worn on the index finger (albeit a somewhat large one), to appear less conspicuous. In addition, we consider the following to be important design considerations: *leveraging the pointing gesture* and *minimal instrumentation of the environment.*

3.3.1.1 Leveraging The Pointing Gesture:

Pointing the index finger at something is a natural and universal deictic gesture [25] used to refer to an object and ask for or convey information about that object [138; 27]. EyeRing augments this natural behavior without obstructing it. We propose to leverage attributes of the pointing gesture, the focus of attention and the implied dialog, as guides for the device to support the interaction.

3.3.1.2 Minimal Instrumentation:

Many input devices require the use of special purpose sensors, and they often only work in an instrumented environment (with markers, beacons, external sensors, etc.) [179; 238; 142]. A design that requires minimal instrumentation of the environment results in a more generic input device. However, a generic system requires additional features to make the interaction specific and focused, which we try to achieve in EyeRing by leveraging the pointing gesture to infer what the user is interested in.

3.3.2 Designing Devices For Visually Impaired Users

The design of EyeRing was informed by recent insights into the design of assistive technology for the visually impaired. The insights made by Shinohara through ethnographic research of assistive technologies struck the right chord with our intentions, most specifically their notions of *independence, portability and social acceptance.* "It is important to consider design ideas supporting cohesive socialization with .. people

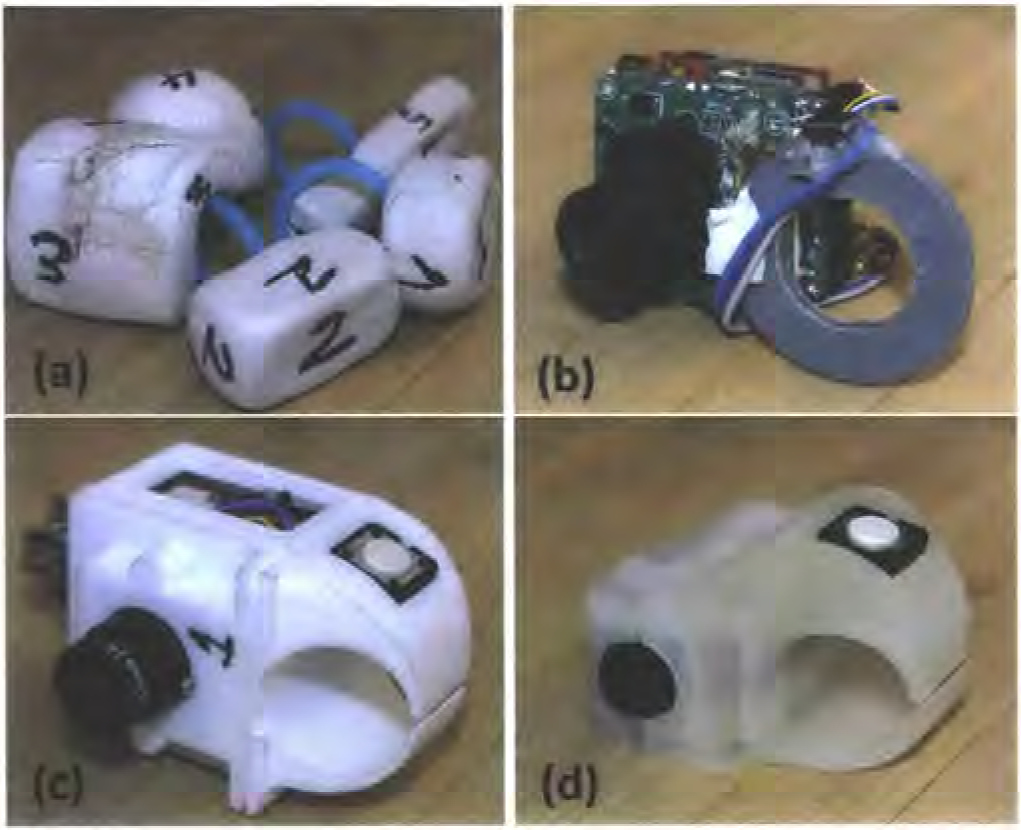


Figure 3-2: Prototypes of the EyeRing. (a) Clay models used for initial exploration. (b) First working prototype. (c) First prototype with plastic casing. (d) Second prototype. The casing for prototypes (c) and (d) was designed by Amit Zoran.

in [a] social sphere," Shinohara claims [209] and goes on to say that "socially acceptable design might draw less unnecessary attention and change misperceptions about assistive devices." This is reiterated in Winberg's work on collaboration using assistive technology: "Non-visual interfaces should .. be coherent with visual interfaces to enable collaboration" [250]. EyeRing is designed to support operation by both sighted and visually impaired users in the same fashion. It is worn on a finger and still allows using the hand and finger for feeling and holding. Even though it was not our primary concern, we also strive to make the device appealing for sighted people to wear, which inherently aligns with the claim that generic devices used both by sighted and non-sighted are far more successful [210].

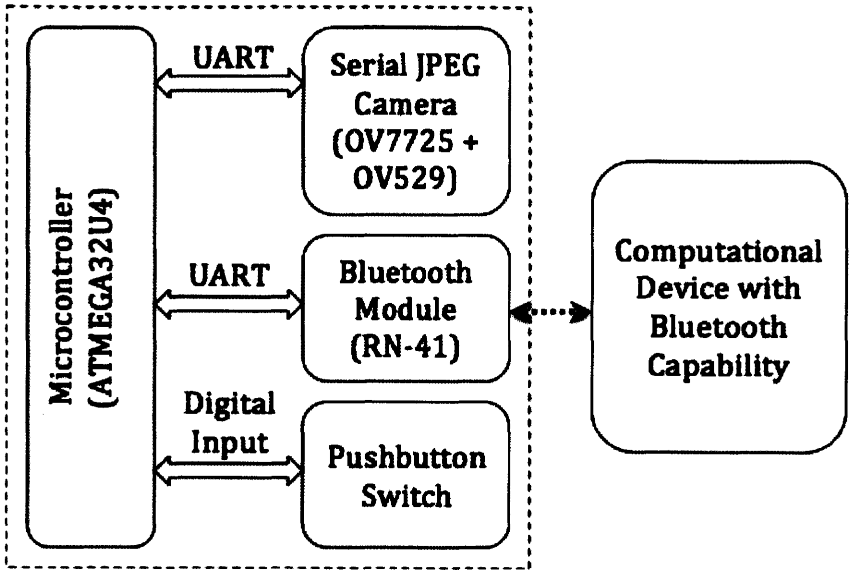


Figure 3-3: Overview of the EyeRing System.

Other design principles that resonated with us were *independence, portability* and *distinguishability of similars.* Applications we propose for visually impaired users are intended to increase self efficacy for blind users, and the small form factor of the ring ensures portability. By distinguishability of similarities Shinohara means "[ability] to distinguish among item with similar features" [209], which is why we focused on implementation of object recognition capabilities.

3.4 Eyering

The EyeRing consists of a finger-worn device with an embedded camera and a computation element - a smartphone or computer, which is also used for speech I/O. The finger-worn device is autonomous, wireless, and includes a single button to initiate the interaction. Information from the device is transferred via Bluetooth to the computing element where it is processed. An overview of the EyeRing system is shown in Figure 3-3.

3.4.1 Hardware Design

The first working prototype of the EyeRing used a JPEG Camera, AVR processor, Bluetooth module, polymer Lithium-ion battery, and push button switch. These components were attached onto a ring-shaped plastic piece (Figure 3-2b). This early working prototype enabled us to explore various application scenarios. Based on preliminary user reactions, we found a need to optimize the design, especially in terms of faster image acquisition and a smaller packaging. As a result, we came up with an improved hardware design for the second EyeRing prototype, which is discussed in detail below (Figure 3-2d).

3.4.1.1 Microcontroller

We chose to use an Atmel 8 bit AVR (ATmega32U4) microcontroller because the EyeRing only requires basic peripherals like digital input/output (I/O) and UART communication. A digital I/O pin is configured as an input and connected to a push button switch for user interaction. Two UARTs are used in the AVR. One is used for serial communication with an image acquisition module, and the other is used for setting up a Bluetooth communication channel.

3.4.1.2 Image acquisition module

The EyeRing design uses an image acquisition module based on the OV7725 VGA CMOS sensor and the OV529 JPEG engine. It uses UART protocol to communicate with a microcontroller for setting up image properties and grabbing image data.

3.4.1.3 Wireless module

In our design consideration, we require an always available, high speed wireless communication protocol. Bluetooth provides a good balance in terms of speed and availability compared to Wifi or Near Field Communication (NFC). Thus, the wireless communication between the EyeRing and a mobile device is established using a Roving Networks RN-42 Bluetooth module with a baud rate of 115 kbps. This translates, under optimal conditions, to approximately 6 JPEG compressed images (with pixel resolution of 320 x 240) per second.

3.4.2 Software Design

We developed a Bluetooth communication module that connects the EyeRing with a smartphone running Android 2.2 or alternatively with a Notebook computer running Windows 7. These modules receive binary image data from the ring, as well as button click events. Some of the computer vision algorithms (e.g. currency recognition, tag recognition) were developed in-house, and we used a 3rd party software [154] for general object recognition. At this point some of our computer vision software runs on the smartphone, while other apps run on the PC depending on the task at hand. The software architecture (communication module and vision engine) allows for easy development of different applications.

3.4.3 Interaction Flow

When used for the first time, EyeRing must be paired with the smartphone or PC application; however, this is done only once and henceforth a Bluetooth connection will be automatically established. Bearing in mind that the device should support both sighted and visually impaired users, we completely rely on non-visual interaction for all usage of the system. A typical interaction starts when the user performs a single click on the pushbutton switch located on the side of the ring using his or her thumb (Figure 3-2c, 3-2d). The type of analysis and corresponding response that follows depend on the selected application (currency, tag, insert, etc.) The user may change to a different application by double clicking the pushbutton and giving the system a brief verbal command that names the application, for example 'insert' (to insert some pictures previously taken into some online document), 'currency' (to recognize the value of a dollar bill), 'tag' (to recognize a price tag), and so on. The applications use a text-to-speech engine to provide audio feedback, hence providing a less disruptive interaction for sighted and visually impaired people alike.

3.5 Eyering Enabled Applications

The EyeRing system opens up the potential to build a great number of applications for people with vision impairments as well as for sighted people. In the following sections, we present a detailed description of two proof-of-concept application scenarios: (1) a shopping assistant, to increase the independence of visually impaired persons in a shopping scenario; and (2) a desktop application providing a seamless copy-paste interaction for sighted people. On top of those, we experimented with a number of additional applications and use cases. One is an interactive application for children in a pre-reading stage that supports situated learning, by letting them read text on their own, before they can recognize alphabets or words. It has been shown that pointing at words while reading them aloud helps children learn faster [178]. When a text consists of many words the EyeRing system assumes that the word they want read is the one at the tip of their finger. Another application currently in development builds upon the idea of I/O brush [190], where the ring is used as a 'paint brush' to capture a texture (for brush stroke) and to draw or paint around a screen or projected canvas. Many of these types of applications may exist for iPads, iPhones or similar devices, however the advantage of the EyeRing is that it makes them instantaneous, requiring minimal effort and reducing shift of attention.

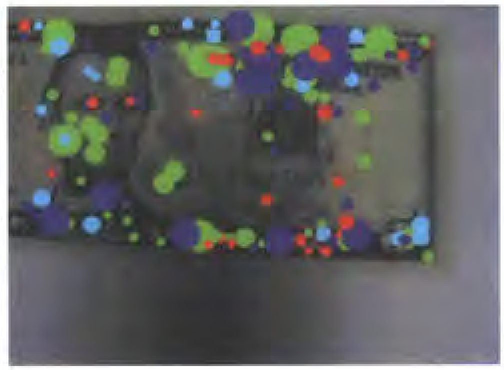
3.5.1 A Shopping Assistant For Blind Users

3.5.1.1 CurrencyDetector

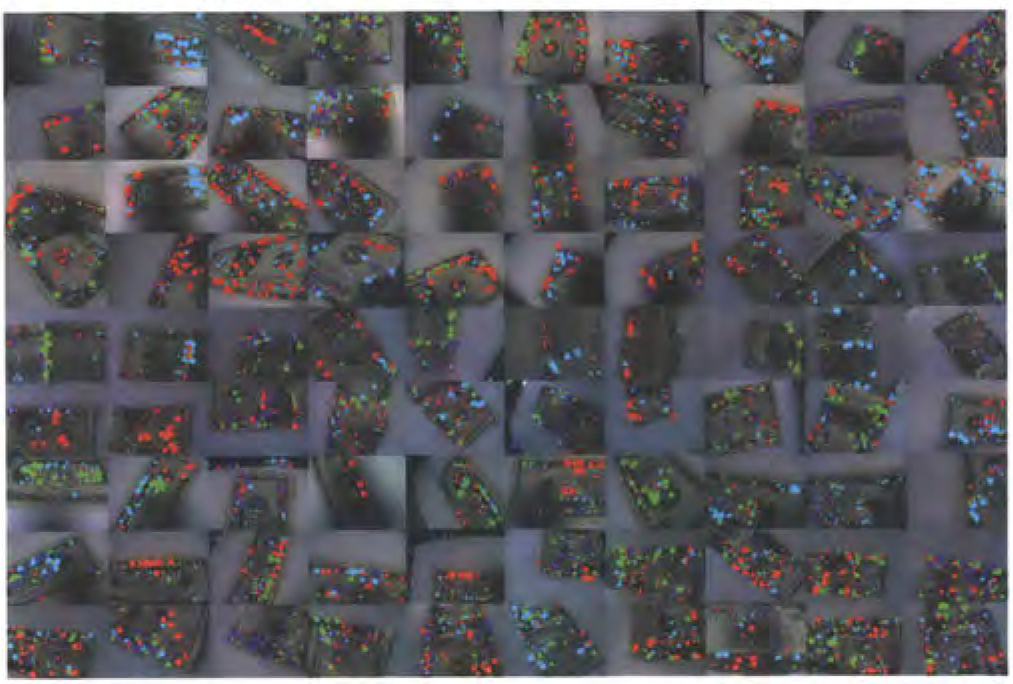
Although currency detection applications for blind users already exist for smartphones [230], these applications require many steps to operate. Specifically, the user has to find the phone, unlock the screen, browse (using a sequential and hence slow auditory approach) to the right app, open the app, take a picture, listen for the answer, turn the phone off and put it away. In contrast, EyeRing requires a fewer number of steps, simply pointing to a currency note and clicking the button, while the other hand is free to hold the note. The system generates synthetic speech output indicating the monetary value of the note. If the currency is not recognized, an error message asks the user to take another picture.

Our EyeRing currency detector application is intended to help a user to identify USA currency bills ($1, $5, $10, $20, $100), although it is easily extendable to other currencies. A detection algorithm based on a Bag of Visual Words (BoVW) approach [1] makes a prediction on the type of note from the input image. Initially, the vocabulary was trained to be 1000 features long and then reduced by arrtribute selection to 170 features. A multi-class SVM (with RBF kernel) was used for classifying. The training dataset consists of 800 images under different lighting conditions and distances, 100 samples were held out for parameter tuning, and the rest were used in a 10-fold cross-validation scheme. For testing, an additional 270 images were used. The overall recognition rate is roughly 92% with a 0.905 kappa statistic. Figure 3-4 illustrates the currency detection process.

(a) Image



(b) Keypoints



(c) Test-Set

Figure 3-4: CurrencyDetector process: (a) Scanned image. (b) Detected features. (c) Montage image showing our test set with detected features - different colors represent different classes.

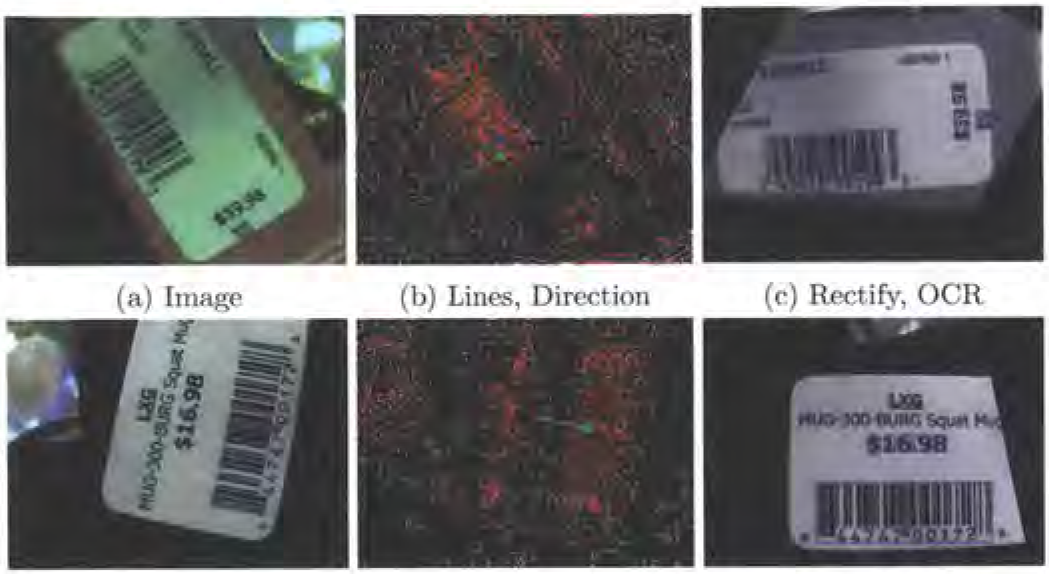
3.5.1.2 TagDetector

The tag detector application intends to assist people with vision impairments in reading price tags on store products. It is based on searching the input image for cues of tag existence, and then extracting textual features or barcodes that allow for retrieving the price. Other applications such as the popular barcode scanner apps (similar to currency detection apps) on smartphones offer the same capabilities, however they too require the user to go through significantly more steps. Using the pointing gesture, we provide a more natural way of aiming the camera at the price tag and getting the result with a single-click operation.

Most product price tags include a UPC-type barcode, as it is a worldwide standard, and the price is usually indicated in a parallel or orthogonal alignment in relation to it (Figure 3-5a, 3-5d). We developed an automatic method to detect the orientation of the barcode in the image to extract the indicated price. A number of researchers recently used edge images and Hough Transform to locate barcodes in images [265; 262], therefore a similar approach was chosen. First a combination of 2^{nd}-order Sobel filters were performed, and then a probabilistic Hough line transform. A line-direction histogram is calculated for each cell of a regular grid over the image. The histogram has 32 bins for line angles, and the grid is of 10x10 cells. A scoring scheme is used to rank the cells:

$$Score(i,j) = \frac{\max_{\theta} Bin_{i,j}(\theta)}{\sum_{Bin_{i,j}(\theta)>0} 1}$$

Where $Bin_{i,j}$ is the angles histogram vector, containing for each angle θ the number of lines in cell i, j agreeing with that angle. The score therefore highly ranks a cell with maximum agreeing lines and minimum possible angles. Unless there is significant interference, this corresponds to a lines-barcode with high probability in our test-set (Figure 3-5b, 3-5e). If fewer than 5 agreeing lines are found, the image is deemed not to contain a barcode. Finally the image is rotated about the center in 4 possible



(a) Image (b) Lines, Direction (c) Rectify, OCR

(d) Image (e) Lines, Direction (f) Rectify, OCR

Figure 3-5: TagDetector process.

rotations.

The next step is Optical Character Recognition (OCR) on the rectified images (Figure 3-5c, 3-5f) to recover any trace of a price mark. A '$' sign in the recovered text serves as an indicator, assuming the price is written to its right. If a price is found, it is spoken to the user, else an error message of either "No Barcode Found" or "No Price Extracted" is played back.

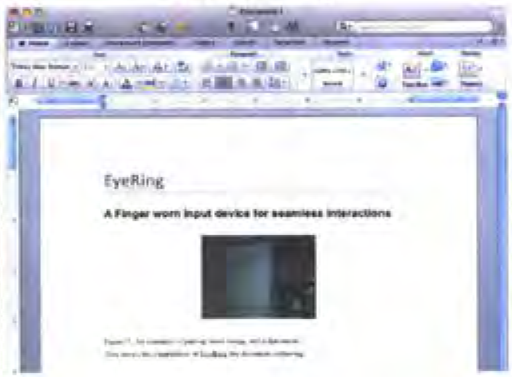
3.5.1.3 Initial Reactions From Visually Impaired Users

During initial development, we brought an EyeRing prototype to a group of people with vision impairments, who are particularly interested in assistive technology. They were given the EyeRing prototype and told how they could use it. Some mentioned a few challenges they have when outside of a well-known environment: navigation, object recognition and reading printed text (which aids both in navigation and recognition) as the tasks which they need most help with. The idea of the shopping assistant

(a) Copying

(b) Pointing

(c) Features detection



(d) Pasting

Figure 3-6: CopyPaster process.

was appealing to them because it helps distinguish objects. However, they raised a few concerns such as hesitation of using a camera as they have little to no experience taking photos, and using the ring with basic phones or different operating systems. When asked to comment on the EyeRing interactions, many users commented that "It (EyeRing) can't get any easier." We observed that the pointing gesture was intuitive to use; however a continuous feedback to assist in pointing at the right object might make it even better. The next chapters in this document describe our process of incorporating continuous-auditory feedback into an augmented finger-pointing device. We also conducted a shopping scenario case study of EyeRing with a blind user who had never used the device before. He easily adapted to using EyeRing and managed to select the 'Cheez-Its' crackers box among many other morphologically similar products.

3.5.2 A Desktop Assistant For Sighted Users

3.5.2.1 CopyPaster

Embedding an image of a proximal object into a digital document commonly involves numerous devices and operations: a camera, smartphone or scanner for capturing, then email, USB drive or a cloud service for transferring, finding the correct place in the document, and lastly using the word processor's commands for embedding. The EyeRing offers a simplified interaction to achieve all of the above steps with minimal effort and direct action. A user would simply point his or her EyeRing to capture any image or text in their environment, directly navigate it on the screen to the required position by pointing again, and paste that information into a document authoring application. The user may optionally scale and rotate the data to the desired transformation, directly by rotating and moving the finger towards the screen. A button click commits the image to the document, similar to the ubiquitous copy-paste operation.

We implemented the interaction by continuously matching SURF features from the camera and those of the screen. A homography relationship is estimated via a robust estimator, and this allows for understanding the position, scale and rotation of the image. For improving the tracking of features on the screen, as well as giving visual feedback, a memory-resident application projects a semi-transparent pattern on the screen at the time of pasting, as well as a 'ghost' of the pasted image. See Figure 3-6 for an illustration.

3.5.3 Exploration of Music Sheet Reading

We began an exploration into reading a different symbolic language than text (of the TagDetector application) - stave musical notation. An image processing pipeline,

Figure 3-7: The sheet music reading system. A user points his finger at a note and hears it played out and displayed on screen

which included detecting the staff lines, locating the note and determining its pitch. A user points the device at a printed sheet of music, pointing the device in the direction of a note and clicks the button to take a picture. The smartphone then analyzes the image and plays the note out, as well as showing it on a small piano-like keyboard on the screen (see Figure 3-7). The system is only able to read a single note at a time in each picture.

3.6 EyeRing vs SmartPhone Experiment

The EyeRing is designed to be an 'immediate' interface. In other words, it should require a minimal number of steps to accomplish a task compared to a smartphone, where a user would have to browse to an app, launch it, point the camera, etc. As a result, the cost-benefit ratio of this input device is better. Thus we hypothesize that the EyeRing may be a faster device to identify single objects, even for sighted people. We conducted a study to compare the EyeRing currency recognizer against a state-of-the-art smartphone application, LookTel [230].

3.6.0.1 Participants and Apparatus

Twelve sighted participants (9 male subjects and 3 female subjects) took part in the study. Their median age was 22 years ranging from 17 to 34 years. EyeRing and a smartphone (iPhone 4S) were used to conduct the study.

3.6.0.2 Procedure

The experiment was a 2×5 within-subjects factorial design. The two independent variables were: device type (EyeRing or smartphone) and number of currency notes (set of 1, 2, 3, 4 and 5). Participants were given a device (EyeRing or smartphone) and asked to identify currency notes ($ 1, 5, 10, 20, 100). The currency notes were given in 5 sets: {$1}, {$1, $5}, {$1, $5, $10}, {$1, $5, $10, $20} and {$1, $5, $10, $20, $100}, presented randomly. Half of the participants used EyeRing followed by the smartphone, the other half used the smartphone followed by EyeRing. With EyeRing, participants pointed at a currency note and pressed the trigger button; subsequently they heard audio feedback (value of the note if identified, otherwise an error message asking for a re-take). For the smartphone application, participants opened the application and scanned using the built-in camera on the phone until they heard the value of the note. The smartphone application had a continuous scanning feature meaning that once started, the app constantly looked for bills and reported their value [230]. For each set of notes, we measured the task completion time (i.e. from the time they were given the set of notes to the time they had identified all the

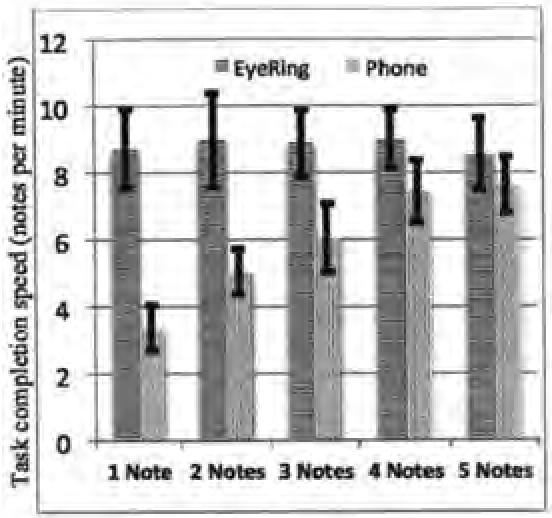


Figure 3-8: Task completion speed (notes per minute) across all experimental conditions (error bars show 95% confidence interval).

notes). All the participants did a practice session to familiarize themselves with the devices and the experiment procedure. After the study, participants were asked to rate their experience by answering a questionnaire. Each participant took approximately 15 minutes to complete the study.

3.6.0.3 Results and analysis

Based on the time taken to complete the note recognition task, we calculated the speed (in notes per second) for all the different experimental conditions. As seen from Figure 3-8, it appears that participants were generally able to compete the task faster with EyeRing. A two-way repeated measures ANOVA analysis showed that there is a main effect of the 'device type' on speed of the task completion ($F(1,110) = 76.08, p < 0.001$). Also, there is a main effect of 'number of notes to detect' on speed of task completion ($F(4,110) = 5.68, p < 0.001$). Moreover, there is an interaction between 'device type' and the 'number of notes to detect' ($F(1,110) = 5.64, p < 0.001$). This combined with the results of Figure 3-8 implies that EyeRing is faster than the smartphone application as long as there are three or fewer notes to detect. In other words, there is an overhead for using the smartphone app (having to browse to the application, open the application, etc). This overhead contributes significantly to the total task time when there are only a few notes to detect. When there are four or more notes, the overhead cost is compensated due to the continuous scanning function of the smartphone application. In reality this overhead is likely to be even greater than what we measured (since the user would have to find the phone, take it out, unlock the screen, etc). In contrast, for EyeRing, the speed of note detection doesn't depend on the number of notes. This is because there is no overhead, and for each note, the participants need to point-and-shoot to identify the value of the note.

3.6.1 Response to the Questionnaire

Participants rated their experience using eight questions on a scale of 0 (strongly disagree) to 5 (strongly agree). Figure 3-9 shows the questions and the summary of responses.

Although many people liked the smartphone form factor (question **a**), most of them indicated that the phone application requires more effort (question **c**). The difference in scores for questions **c** and **d** is statistically significant. This suggests that EyeRing required less effort compared to a smartphone application. Most participants agreed that the EyeRing offers hands-free operation (question **e**) and that the pointing gesture is helpful to frame the picture (questions **i**). This is in line with our observation that participants touched the note with their finger and then backed off a bit to take a picture. The fact that the EyeRing is more of a pointing device helped them to do the task more easily; however, a couple of participants mentioned that they prefer to get feedback about what they are pointing at. We are currently exploring options for doing so, for example, by using a laser pointer to indicate the location of the camera focus. In contrast, two participants mentioned that the continuous scanning with

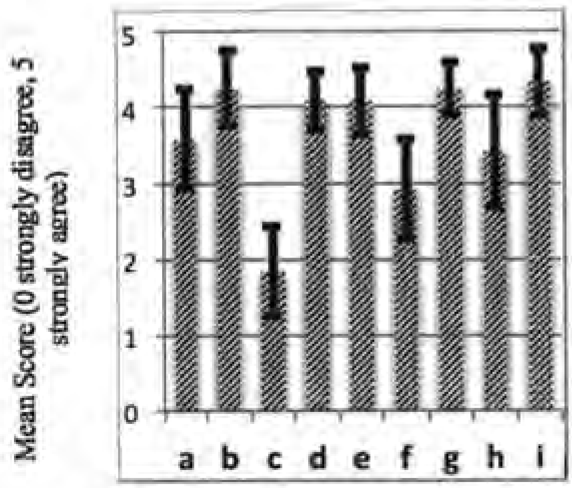


Figure 3-9: Summary of responses to the survey question (error bars show 95% confidence interval). (a) I like the smartphone form factor; (b) I like the EyeRing form factor; (c) Phone app required less effort compared to the EyeRing; (d) the EyeRing required less effort compared to Phone application; (e) the EyeRing allowed hands-free operation; (f) I prefer to use smartphone application on a regular basis; (g) I prefer to use the EyeRing on a regular basis; (h) Camera preview of the smartphone was helpful to frame the picture; (i) Pointing gesture of the EyeRing was helpful to frame the picture.

smartphone made the task easier. In summary, we believe that decoupling the camera from a smartphone made the EyeRing into an input device that is immediately accessible, requires less effort and is less disruptive.

3.7 Conclusion

The EyeRing, which leverages the pointing gesture, shows the potential of building seamless interactions for people with vision impairments and sighted people alike. The nature and design of the ring apparatus is driven by lessons from established design frameworks for both natural interaction and assistive technologies. Preliminary user reactions suggested that the use of EyeRing applications is intuitive and seamless. A controlled user study indicated that EyeRing is an immediately accessible device that performs faster than a smartphone application for a single object detection task.

The following chapter, describing the first version of the FingerReader device, reveals how the EyeRing evolved in terms of form factor, application and target audience. The central lessons learned from the EyeRing that informed the continuing work were: (1) adding a continuous feedback will support a real time operation, (2) using a high frame-rate imaging module will enable additional applications, and (3) the pointing gesture can scaffold tasks of touching and non-visual navigation on an object. In the FingerReader we implemented some of these improvements and tested them with visually impaired persons.

Chapter 4

The FingerReader

A Finger Wearable Device For Reading Text On-The-Go

Accessing printed text in a mobile context is a major challenge for blind individuals. A preliminary study with blind people reveals numerous difficulties with existing state-of-the-art technologies including problems with alignment, focus, accuracy, mobility and efficiency. In this chapter, we present a finger-worn device, FingerReader, that assists blind users with reading printed text on the go. We introduce a novel computer vision algorithm for local-sequential text scanning that enables reading single lines, blocks of text or skimming the text with complementary, multimodal feedback. This system is implemented in a small finger-worn form factor, that enables a more manageable eyes-free operation with trivial setup. We offer findings from three studies performed to determine the usability of the FingerReader.

4.1 Introduction

Some people with a visual impairment (VI) find it difficult to access text documents in different situations, such as reading text on the go and accessing text in non-ideal conditions (e.g. low lighting, unique layout, non-perpendicular page orientations), as reported in interviews we conducted with assistive technology users. We found that available technologies, such as smartphone applications, screen readers, flatbed scanners, e-Book readers, and embossers, are considered to have slow processing speeds, poor accuracy or cumbersome usability. Day-to-day text-based information, such as bus and train station information boards, are said to be generally inaccessible, which greatly affects the mobility and freedom of people with a VI outside the home, the Royal National Institute of Blind People (RNIB) reports [164]. Technological barriers inhibit blind people's abilities to gain more independence, a characteristic widely identified as important by our interviewees.

This chapter presents our work on creating a mobile device to tackle some of the problems current text reading technologies present to blind users. Our work contributes to the growing pool of assistive reading devices in three primary ways:

- First, we share the results of interview sessions with blind users that uncover problems with existing text reading solutions, as well as expectations for future assistive devices and their capabilities. Our design choices are based on these findings.
- Second, we conceptualize and implement FingerReader, a finger-worn system for local-sequential text scanning, where the user scans the text progressively in a local view and hears the recognized words synthesized to audible speech. It enables continuous feedback to the user and allows for new ways of reading, such as non-linear skimming to different parts of the text. Our proposed

method utilizes computer vision algorithms, along with audio and tactile cues for effectively guiding the user in reading printed text using the fingertip as a cursor.

- Last, we report findings from three evaluations: a technical evaluation to understand the text extraction accuracy, user feedback sessions with blind participants to assess the feedback mechanism, and an end-to-end study to assess the system's real-world feasibility and explore further design opportunities.

4.2 Related Work

Researchers in both academia and industry exhibited a keen interest in aiding people with VI to read printed text. The earliest evidence we found for a specialized assistive text-reading device for the blind is the Optophone, dating back to 1914 [39]. However the Optacon [123], a steerable miniature camera that controls a tactile display, is a more widely known device from the mid 20th century. Table 4.1 presents more contemporary methods of text-reading for the VI based on key features: adaptation for non-perfect imaging, type of text, User Interface (UI) suitable for VI and the evaluation method. Thereafter we discuss related work in three categories: wearable devices, handheld devices and readily available products.

Publication	Year	Interface	Type of Text	Response Time	Adaptation	Evaluation	Accuracy
Ezaki et al. [43]	2004	PDA	Signage			ICDAR 2003	P 0.56 R 0.70
Mattar et al. [135]	2005	Head-worn	Signage		Color, Clutter	Dataset	P ?.?? R 0.90[1]
Hanif and Prevost [61]	2007	Glasses, Tactile	Signage	43-196s		ICDAR 2003	P 0.71 R 0.64
SYPOLE [167]	2007	PDA	Products, Book cover	10-30s	Warping, Lighting	VI users	P 0.98 R 0.90[1]
Pazio et al. [165]	2007		Signage		Slanted text	ICDAR 2003	
Yi and Tian [261]	2012	Glasses	Signage, Products	1.5s	Coloring	VI users	P 0.68 R 0.54
Shen and Coughlan [202]	2012	PDA, Tactile	Signage	$<$1s		VI users	
Kane et al. [89]	2013	Stationery	Printed page	Interactive	Warping	VI users	
Stearns et al. [227]	2014	Finger-worn	Printed page	Interactive	Warping	VI users	
Shilkrot et al. [204]	2014	Finger-worn	Printed page	Interactive	Slanting, Lighting	VI users	

[1] This report is of the OCR / text extraction engine alone and not the complete system.

Table 4.1: Recent efforts in academia of text-reading solutions for the VI. Accuracy is in precision (P) recall (R) form, as reported by the authors.

The reader is invited to read Chapter 2 for an overview of much of the related finger worn devices to the FingerReader. Additionally, the encompassing survey by Lévesque [118] provides insight into the use of tactile feedback in assistive technology.

4.2.1 Wearable Devices

In a wearable form-factor, it is possible to use the body as a directing and focusing mechanism, relying on proprioception or the sense of touch, which are of utmost importance for people with VI. Yi and Tian [261] placed a camera on shade-glasses to recognize and synthesize text written on objects in front of them, and Hanif and Prevost's [61] did the same while adding a handheld device for tactile cues. Mattar et al. are using a head-worn camera [135], while Ezaki et al. developed a shoulder-mountable camera paired with a PDA [43]. Differing from these systems, we proposed using the finger as a guide and supporting sequential acquisition of text rather than reading text blocks (in a preliminary publication [204]). This concept has inspired other researchers in the community, such as Stearnes et al. [227].

4.2.2 Handheld and Mobile Devices

Mancas-Thillou, Gaudissart, Peters and Ferreira's SYPOLE consisted of a camera phone/PDA to recognize banknotes, barcodes and labels on various objects [167], and Shen and Coughlan recently presented a smartphone based sign reader that incorporates tactile vibration cues to help keep the text-region aligned [202]. The VizWiz mobile assistive application takes a different approach by offloading the computation to humans, although it enables far more complex features than simply reading text, it lacks real time response [19].

4.2.3 Assistive Mobile Text Reading Products

Mobile phone devices are very prolific in the community of blind users for their availability, connectivity and assistive operation modes, therefore many applications were built on top of them: the kNFB kReader[1], Blindsight's Text Detective[2], ABBYY's Text Grabber[3], StandScan[4], SayText[5], ZoomReader[6] and Prizmo[7]. Meijer's vOICe for Android project is an algorithm that translates a scene to sound; recently they introduced OCR capabilities and enabling usage of Google Glass[8]. ABiSee's EyePal ROL is a portable reading device, albeit quite large and heavy[9], to which OrCam's recent assistive eyeglasses[10] or the Intel Reader[11] present a more lightweight alternative.

Prototypes and products in all three categories, save for [227], follow the assumption that the goal is to consume an entire block of text at once, therefore requiring to image the text from a distance or use a special stand. In contrast, we focused on creating a smaller and less conspicuous device, allowing for intimate operation with the finger that will not seem strange to an outside onlooker, following the conclusions of Shinohara and Wobbrock [212]. Giving the option to read locally, skim over the text at will in a varying pace, while still being able to read it through, we sought to create a more liberating reading experience.

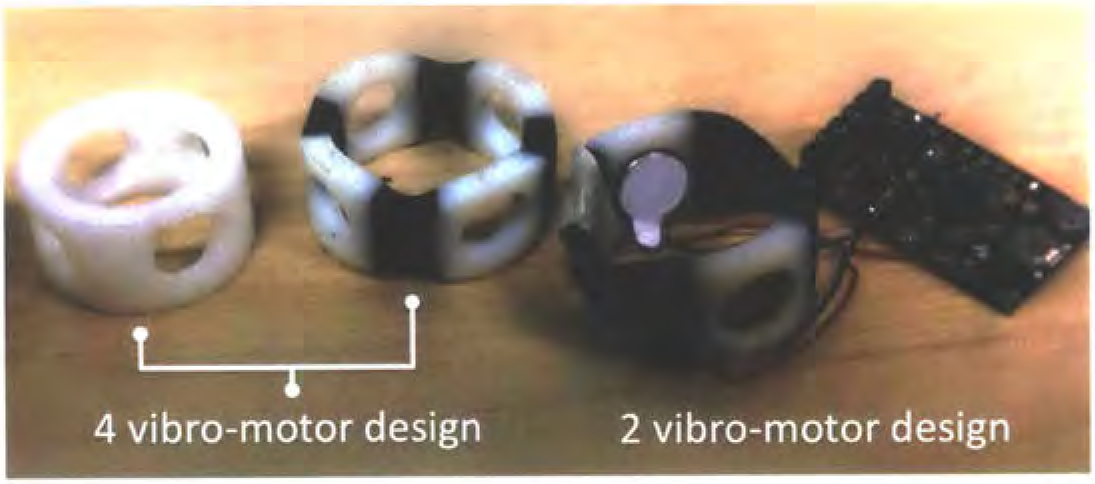
[1] http://www.knfbreader.com
[2] http://blindsight.com
[3] http://www.abbyy.com/textgrabber
[4] http://standscan.com
[5] http://www.docscannerapp.com/saytext
[6] http://mobile.aisquared.com
[7] http://www.creaceed.com/iprizmo
[8] http://www.seeingwithsound.com
[9] http://www.abisee.com
[10] http://www.orcam.com
[11] http://www.intel.com/pressroom/kits/healthcare/reader/

4.3 Focus Group Sessions

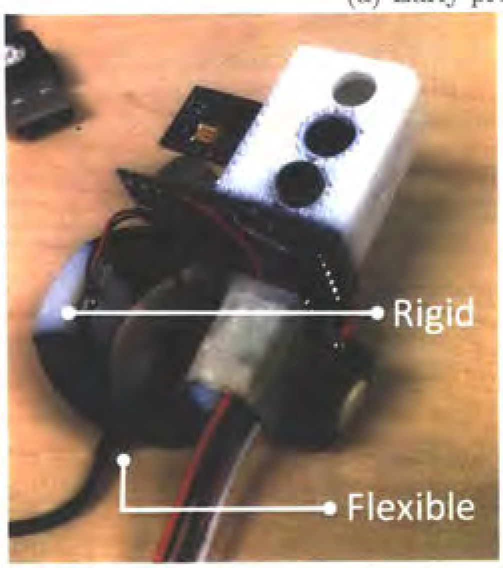
We conducted two sessions with congenitally blind users ($N_1 = 3$, $N_2 = 4$) to gain insights into their text reading habits, and identify concerns with existing technologies. We also presented simple prototypes of the FingerReader (see Figure 4-1a) later in each session to get opinions on the form factor and elicit discussion on the intended usage pattern. The two sessions went on for roughly 5 hours, so only the most relevant findings are summarized herein:

- All participants routinely used flatbed scanners and camera-equipped smartphones to access printed text.
- While flatbed scanners were reported to be easy to use, participants mentioned problems when scanning oddly shaped prints. Our participants preferred mobile devices due to their handiness, but again reported issues with focusing the camera on the print. Overall, both approaches were considered inefficient. One participant went on to say: "*I want to be as efficient as a sighted person*".
- Reported usability issues revolved around text alignment, recognition accuracy, software processing speed, and problems with mitigating low lighting conditions. Information return rates were marked as important, where at times digitizing a letter-sized page could take up to 3 minutes.
- Participants also showed interest in reading fragments of text such as off a restaurant menu, text on screens, business cards, and canned goods labels. A smaller device was also preferred, as well as a single-handed, convenient operation.

Following the findings from the focus group sessions, we set to design a device that enables: skimming through the text, have a real time single-handed operation, and provides multimodal continuous feedback.



(a) Early prototypes Evolution



(b) Multi-material prototype

(c) New prototype

Figure 4-1: FingerReader prototypes.

4.4 FingerReader: A Wearable Reading Device

FingerReader is an index-finger wearable device that supports the blind in reading printed text by scanning with the finger and hearing the words as synthesized speech (see Figure 4-1c). Our work features hardware and software that includes video processing algorithms and multiple output modalities, including tactile and auditory channels.

The design of the FingerReader is a continuation of our work on the EyeRing (Chapter 3), and inspired by additional focus group sessions we performed. Exploring the design concepts with blind users revealed the need to have a small, portable device that supports free movement, requires minimal setup and utilizes real-time, distinctive multimodal response. The finger-worn design keeps the camera in a fixed distance from the text and utilizes the inherent finger's sense of touch when scanning text on the surface. Additionally, the device provides a simple interface for users as it has no buttons, and affords to easily identify the side with the camera lens for proper orientation.

4.4.1 Hardware Details

The FingerReader hardware features tactile feedback via vibration motors, a dual-material case design inspired by the focus group sessions and a high-resolution mini video camera. Vibration motors are embedded in the ring to provide tactile feedback on which direction the user should move the camera via distinctive signals. Initially, two ring designs were explored: 4 motor and 2 motor (see Fig. 4-1a). Early tests with blind users showed that in the 2 motor design signals were far easier to distinguish than with the 4 motor design, as the 4 motors were too close together. This led to a new, multi-material design using a white resin-based material to make up the

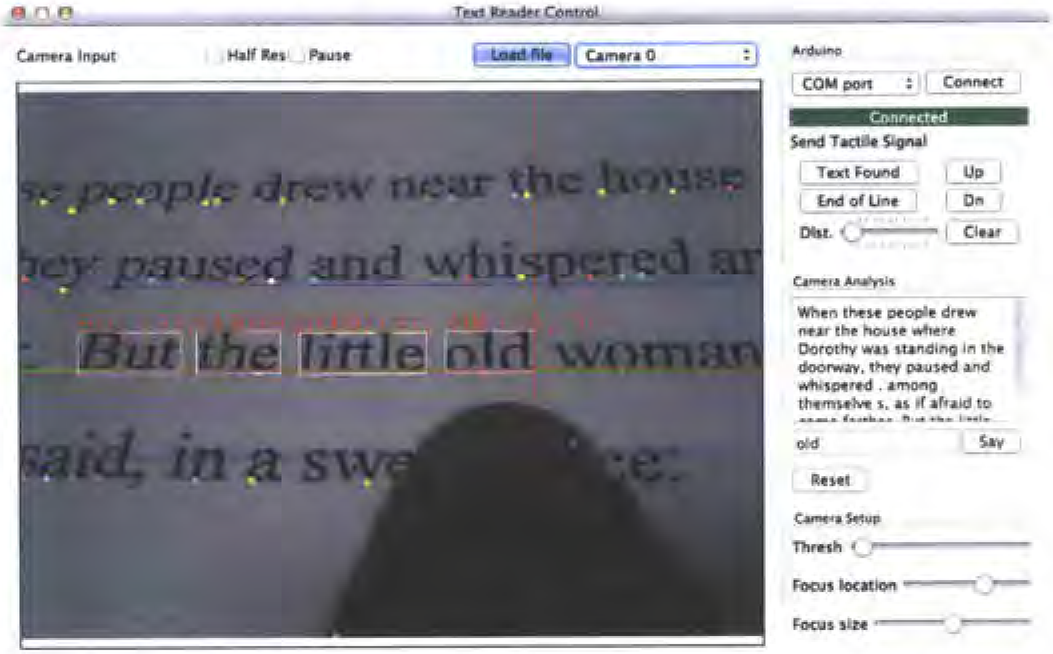


Figure 4-2: Our software in midst of reading, showing the detected line, words and the accumulated extracted text

harder sections where the motors are embedded and a rubbery material for the flexible connections. The dual material design provides flexibility to the ring's fit as well as helps dampen the vibrations and reduce confusion for the user.

4.4.2 Algorithms and Software

We developed a software stack that includes a sequential text reading algorithm, hardware control driver, integration layer with Tesseract OCR [220] and Flite Text-to-Speech (TTS) [22], currently in a standalone PC application (see Fig. 4-2).

4.4.2.1 Vision Algorithm Overview

The sequential text reading algorithm is comprised of a number of sub-algorithms concatenated in a state-machine (see Fig. 4-3), to accommodate for a continuous operation by a blind person. The first two states (Detect Scene and Learn Finger) are used for calibration for the higher level text extraction and tracking work states (No Line, Line Found and End of Line). Each state delivers timely audio cues to the users to inform them of the process. All states and their underlying algorithms are detailed in the following sections.

The operation begins with detecting if the camera indeed is looking at a close-up view of a finger touching a contrasting paper, which is what the system expects in a typical operation. Once achieving a stable view, the system looks to locate the fingertip as a cursor for finding characters, words and lines. The next three states deal with finding and maintaining the working line and reading words. For finding a line, the first line or otherwise, a user may scan the page (in No Line mode) until receiving an audio cue that text has been found. While a text line is maintained, the system will stay in the Line Found state, until the user advanced to the end of the line or the line is lost (by moving too far up or down from the line or away from the paper).

Scene and Finger Detection: The initial calibration step tries to ascertain whether the camera sees a finger on a contrasting paper. The input camera image is converted to the normalized-RGB space: $(R, G, B) = \left(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b}\right)$, however we keep only the normalized red channel (R) that corresponds well with skin colors and

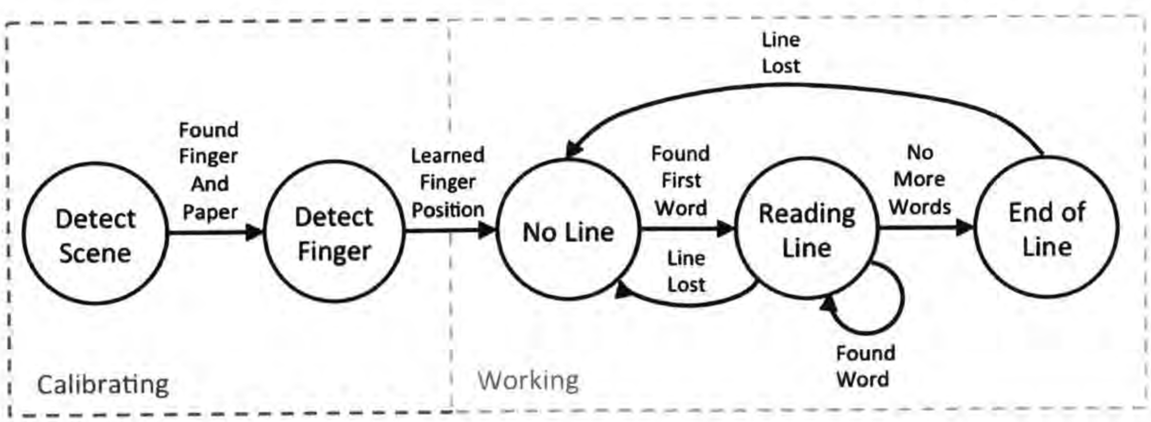


Figure 4-3: Sequential text reading algorithm state machine.

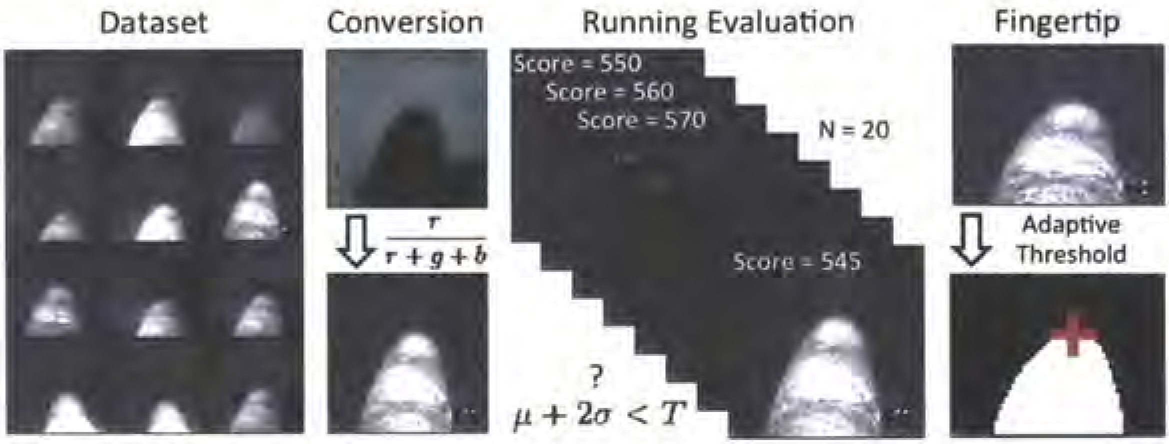


Figure 4-4: Scene and fingertip detection.

ameliorates lighting effects. The monochromatic image is downscaled to 50x50 pixels and matched to a dataset of prerecorded typical images of fingers and papers from a proper perspective of the device camera. To score an incoming example image, we perform a nearest neighbor matching and use the distance to the closest database neighbor. Once a stable low score is achieved (by means of a running-window of 20 samples and testing if $\mu_{score} + 2\sigma < threshold$) the system deems the scene to be a well-placed finger on a paper, issues an audio command and advances the state machine. See Fig. 4-4 for an illustration of this process.

In the finger detection state we binarize the R channel image using Otsu adaptive thresholding and line scan for the top white pixel, which is considered a candidate fingertip point (see Fig. 4-4). During this process the user is instructed not to move, and our system collects samples of the fingertip location from which we extract a normal distribution. In the next working states the fingertip is tracked in the same fashion from the R channel image, however, in this case, we assign each detection with a probability measure based on the learned distribution to eradicate outliers.

The inlying fingertip detection guides a local horizontal *focus region*, located above the fingertip, within which the following states perform their operations. The focus region helps with efficiency in calculation and also reduces confusion for the line

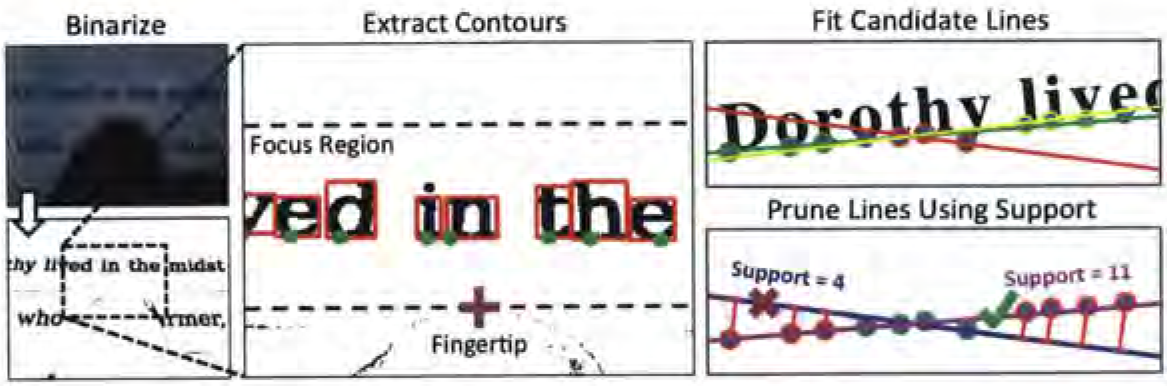


Figure 4-5: Text line extraction process.

extraction algorithm with neighboring lines (see Fig. 4-5). The height of the focus region may be adjusted as a parameter, but the system automatically determines it once a text line is found.

Line Extraction: Within the focus region, we start with local adaptive image binarization (using a shifting window and the mean intensity value) and selective contour extraction based on contour area, with thresholds for typical character size to remove outliers. We pick the bottom point of each contour as the baseline point, allowing some letters, such as 'y','g' or 'j' whose bottom point is below the baseline, to create artifacts that will later be pruned out. Thereafter we look for candidate lines by fitting line equations to triplets of baseline points; we then keep lines with feasible slopes and discard those that do not make sense. We further prune by looking for supporting baseline points to the candidate lines based on distance from the line. Then we eliminate duplicate candidates using a 2D histogram of slope and intercept that converges similar lines together. Lastly, we recount the corroborating baseline points, refine the line equations based on their supporting points and pick the highest scoring line as the detected text line. When ranking the resulting lines, additionally, we consider their distance from the center of the focus region to help cope with small line spacing, when more than one line is in the focus region. See Fig. 4-5 for an illustration.

Word Extraction: Word extraction is performed by the Tesseract OCR engine on image blocks from the detected text line. Since we focus on small and centric image blocks, the effects of homography between the image and the paper planes, and lens distortion (which is prominent in the outskirts of the image) are negligent. However, we do compensate for the rotational component caused by users twisting their finger with respect to the line, which is modeled by the equation of the detected line.

The OCR engine is instructed to only extract a single word, and it returns: the word, the bounding rectangle, and the detection confidence. Words with high confidence are retained, uttered out loud to the user, and further tracked using their bounding rectangle as described in the next section. See Fig. 4-6 for an illustration.

Word Tracking and Signaling: Whenever a new word is recognized it is added to a pool of words to track along with its initial bounding rectangle. For tracking we use template matching, utilizing image patches of the words and an L_2-norm matching score. Every successful tracking, marked by a low matching score and a feasible tracking velocity (i.e. it corresponds with the predicted finger velocity for that frame), contributes to the bank of patches for that word as well as to the prediction of finger velocity for the next tracking cycle. To maintain an efficient tracking, we do not search the entire frame but constrain the search region around the last position of the word while considering the predicted movement speed. We also look out for blurry patches, caused by rapid movement and the camera's rolling shutter, by binarizing

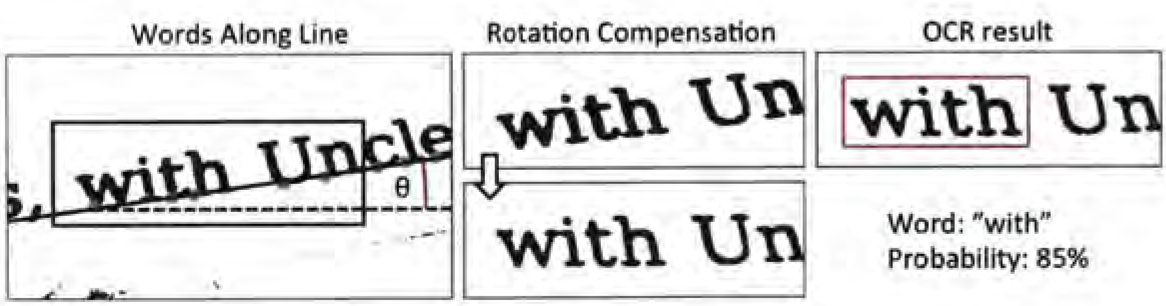


Figure 4-6: Word extraction process.

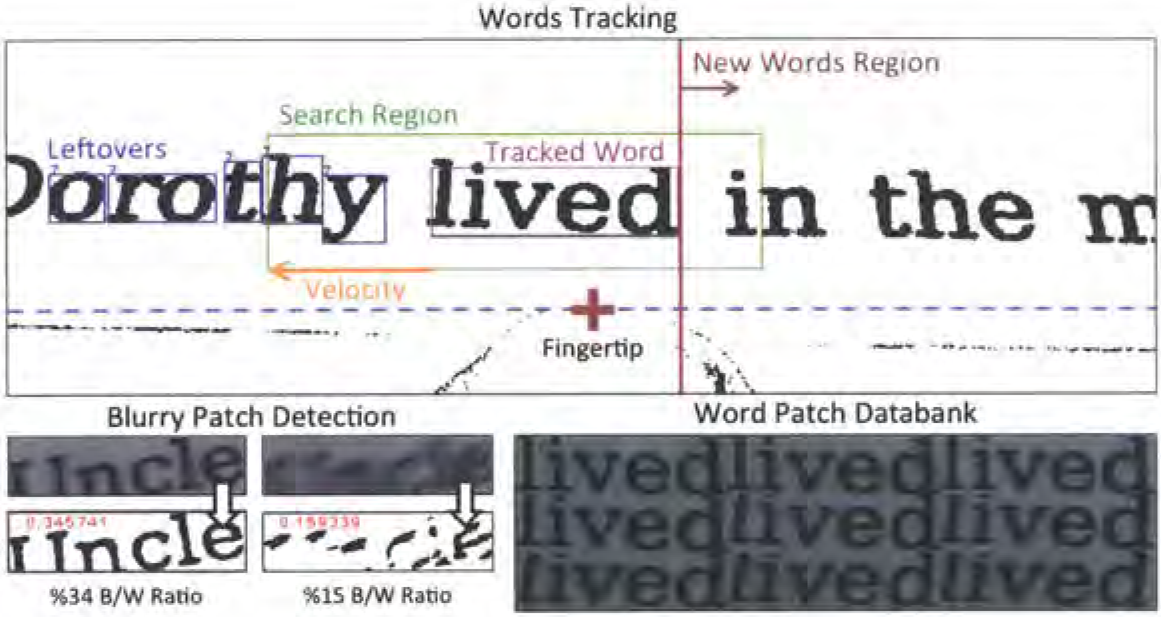


Figure 4-7: Word tracking process.

the patch and counting the number of black vs. white pixels. A ratio of less than 25% black is considered a bad patch to be discarded. If a word was not tracked properly for a set number of frames we deem as "lost", and remove it from the pool. See Fig. 4-7 for an illustration.

We do not dispose of 'lost words' immediately, rather split them to 'leftovers', which are single character patches we track similarly. This way, when a word is phasing out of the frame its remaining characters can still contribute to the prediction of finger speed and the robustness of tracking. When leftovers are not properly tracked they too are discarded.

When the user veers from the scan line, detected using the line equation and the fingertip point, we trigger a gradually increasing tactile and auditory feedback. When the system cannot find more word blocks further along the scan line, it triggers an event and advances to the End of Line state.

Typical frame processing time is less than 20ms, which is suitable for realtime processing. Fast running time is important to enable skimming text as well as for immediate feedback on the scan progress.

4.5 Evaluation

The central question that we sought to explore was how and whether FingerReader can provide effective access to print and reading support for VI users. Towards this end, we conducted a series of evaluations. First, we conducted a technical evaluation to assess whether the FingerReader is sufficiently accurate, and parallelly, user feedback sessions to investigate the usefulness of the different feedback cues with four congenitally blind users. We then used the results from these two fundamental investigations to conduct a qualitative evaluation of FingerReader's text access and reading support with 3 blind users. In the following, we briefly report on the technical analysis and user feedback sessions. We then describe the qualitative evaluation comprehensively and highlight major findings.

Across all studies we were only able to recruit a small number of participants. In accordance with Sears and Hanson, who attest to the difficulty of recruiting participants with similar impairment condition in accessibility research [198]. Thus we looked to maximize the results we can obtain from a small number of participants with different impairment histories instead of striving for generalizability (sensu [211]).

4.5.1 Technical Accuracy Analysis

The main objective of the accuracy analysis was to assess whether the current implementation of the FingerReader is sufficiently accurate to ensure that future user studies will yield unbiased data.

The accuracy was defined as as $acc = 1 - LD_{norm}$, with LD_{norm} being the normalized Levenshtein Distance (LD) [117] between the scanned and original text. The LD counts the number of character edits between two strings, e.g. $LD(\text{“hello”}, \text{“h3ll0”}) = 2$; a higher distance means a less accurate scan. To normalize, we divided the LD by the number of characters in the paragraph (either scanned or original) with the maximal number of characters: $LD_{norm} = LD/\max(S_{scan}, S_{orig})$, where S_i is the length of scanned string i (the scanned string can be larger than the original).

As the test corpus, we randomly chose a set of 65 paragraphs from Baum's “The Wonderful Wizard of Oz”, where each paragraph contained a different number of characters (avg. 365). The book was typeset in Times New Roman, 12pt with 1.5 line spacing.

We measured the accuracy of the text extraction algorithm under optimal conditions (sighted user, adequate lighting) at 93.9% ($\sigma = 0.037$), which verifies that this part of the system works properly. Error analysis shows that most errors occur due to short lines (e.g. of a conversation: “Hello, how are you?” $\rightarrow$ “Hello, how are you? you?”), where FingerReader duplicated the end of the line, therefore increasing the LD. Following this finding, we installed a specific mechanism to prevent words from repeating.

4.5.2 User Feedback on Cueing Modalities

We also conducted user feedback sessions with 4 congenitally blind users to (1) uncover potential problems with the usability of the final design and (2) to compare the usefulness of the feedback. The five feedback types were individually presented, fully counterbalanced: (i) audio, (ii) tactile regular, (iii) tactile fade, (iv) audio and tactile regular, (v) audio and tactile fade. *Tactile fade* produced a gradually increasing vibration (quantized to 10 levels) to indicate vertical deviation from the line, and *tactile* *regular* produced a constant vibration when a certain threshold of deviation from the line was passed. The audio cue was a simple spoken utterance of "up" or "down". After introducing the concepts of using the FingerReader, we used a wooden tablet with a paper displaying a printed paragraph of text to test the four feedback options. A session with a single user went on for roughly 1 hour, included semi-structured interviews, and observation was used for the data gathering method.

The task participants were given was to trace three lines of text using the feedbacks for guidance. We then asked for their preference and impressions on the usability of the device. Analysis of the results showed that participants preferred *tactile fade* compared to other cues (100% preferred *tactile fade*), and recognized the additional information on a gradual deviation from the line. Additionally, tactile fade response provided a continuous feedback, where the other modalities were fragmented. One user reported that "*when [the audio] stops talking, you don't know if it's actually the correct spot because there's no continuous updates, so the vibration guides me much better.*" Our study participants were able to imagine how FingerReader can help them conduct daily tasks, and be able to explore printed text in their surroundings in a novel way.

4.5.3 Print Access and Reading Study

As a next step in the evaluation process, we built upon the prior results and conducted a user study with three blind participants to qualitatively investigate the effectiveness of FingerReader to access and read print. The two main goals of our study were:

1. Analyze the participant's usage of the FingerReader and
2. Investigate the effectiveness of FingerReader for accessing and reading.

We investigated these goals depending on different document types that users will potentially encounter, inspired by findings from prior design probe sessions, and their impairment history, i.e. whether they were congenitally or late blind.

4.5.3.1 Participants and Study Design

Following the approach of Sears and Hanson [198], we hereby detail the participants information. All participants were blind, P2 and P3 since birth and consequently have never experienced text visually (see table 4.2). P1 became blind at the age of 18. Before that, he considered himself an avid reader. P2 has very low light perception, P3 no light perception at all. All participants had perfect hearing and were right-handed. All participants had prior exposure to the FingerReader, which included brief demonstrations during recruitment to make sure participants are comfortable before committing to the study.

They all share stationary text access habits, e.g. in using a screenreader like JAWS to access digital text on a PC or Mac or in scanning printed documents to have them read back e.g. with ABBYY FineReader. On the go, P1 and P2 mostly rely on the help of sighted people to read relevant text to them. Specifically, P2 has never owned a smartphone and does not consider himself tech-savvy. Both P1 and P3 own an iPhone and use it to access digital information using Apple's VoiceOver technology. Yet, P2 considers himself only an occasional user of technology. P3 was the most tech-savvy participant. He regularly uses mobile applications on his iPhone to access printed text on the go, namely TextGrabber and Prizmo. P3 stated that he uses either software as a backup in case the other fails to detect text properly. He described himself as an avid user of a foldable StandScan, yet he seldom carries it with him as it is too 'bulky and cumbersome'. In mobile settings, he usually captures documents free-handedly by applying a two-step process where he first places the print in landscape and centers the iPhone on top of it (*framing*) and then lifts the

	Age	Visual Impairment	Text access habits
P1	27	Blind (since 18)	*Digital:* PC: JAWS, iPhone: VoiceOver *Print:* Volunteer, ABBYY FineReader
P2	53	Light perception (congenital)	*Digital:* PC: JAWS *Print:* Volunteer, flatbed scanner
P3	59	Totally blind (congenital)	*Digital:* PC & iPhone: VoiceOver *Print:* iPhone apps, volunteer, scanner

Table 4.2: Overview of the participants from the text exploration study.

iPhone chin-high to take a picture and have the software read the text back to him (*capture*). The whole capturing process takes him on average 2.5 minutes, excluding any trial and error and without having the text read back to him.

The study took place over two single-user sessions per participant with 3 days in-between sessions to allow the participants to accommodate to the FingerReader technology, have enough time to thoroughly practice with the feedback modalities in both sessions and reflect on their usage. The first session focused on introducing the participants to FingerReader and different document formats. The session lasted 90 minutes in average. The second session focused more on assessing the participants' text access and reading effectiveness, which lasted about 60 minutes in average.

4.5.3.2 Method and Tasks

Fig. 4-9 shows an outline of how the two single-user sessions were run. Each session contained both pre- and post-interviews and practice sessions. We distinguished between two main types of tasks: text *access* and text *reading*. Both types were motivated by insights from the focus group sessions, where participants mentioned that is key for them to simply *access* a printed document to extract their contents (e.g. find the entrees on a restaurant menu) and then zero in on a particular part to *read* its content.

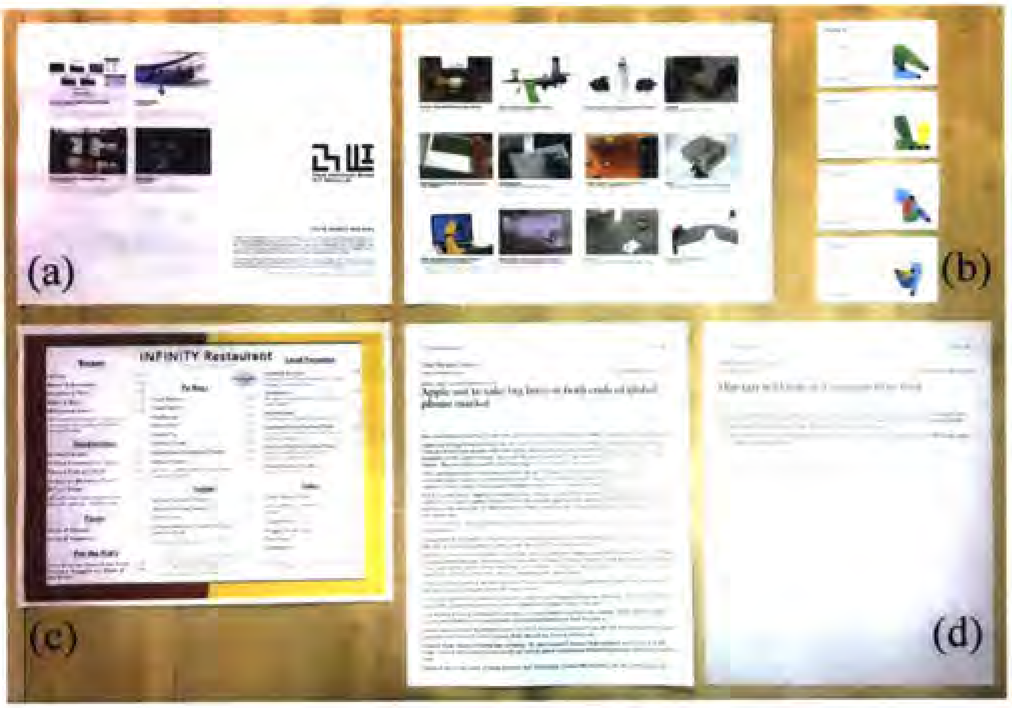


Figure 4-8: Documents used in the study: (a) pamphlet, (b) business cards, (c) restaurant menu and (d) newspaper articles. The font size varied across documents: 10-14pt (business cards, pamphlet), 12-24pt (menu), 12-14pt (newspaper articles).

Session 1: In the first session, each participant was introduced to the core concepts of the FingerReader. Although all participants had prior exposure to the FingerReader, all feedback modalities were explained in detail and an average of 30 minutes were given to practice with the FingerReader on a sample page (a random page from Baum's "The Wonderful Wizard of Oz"). Afterwards, each participant was asked to access three different document types using the FingerReader (see Figure 4-8): (i) a pamphlet with a column layout that also contained pictures, (ii) an A4-sized restaurant menu, three-column layout without pictures and (iii) a set of three business cards, printed in landscape. These document types were inspired by user feedback we obtained in the focus group sessions with design probes, where participants mentioned those documents to be key for them for on-the-go access. The primary task for each participant was to simply use the FingerReader and see whether they can elicit the contents.

Session 2: The second session included a short practice session to let participants refamiliarize themselves with the device. This was followed by a repetition of the text access tasks to qualitatively compare their performance to the first session. Next, each participant was given a set of 5 articles taken from the online edition of a local newspaper (see Fig. 4-8). All articles were set in a single-column layout and did not contain pictures. Each participant was asked to explore the news articles and report the gist of the article. The sessions were concluded with a questionnaire (inspired by [62]).

Each set of tasks was fully counterbalanced, and all feedback modalities were available. As for data gathering techniques, we video-recorded the interactions, lead semi-structured interviews, observed the participants during the session and asked the participants to think aloud. Two authors acted as experimenters in both sessions. One experimenter had perfect vision while the other was blind with only little peripheral light perception (since age 10).

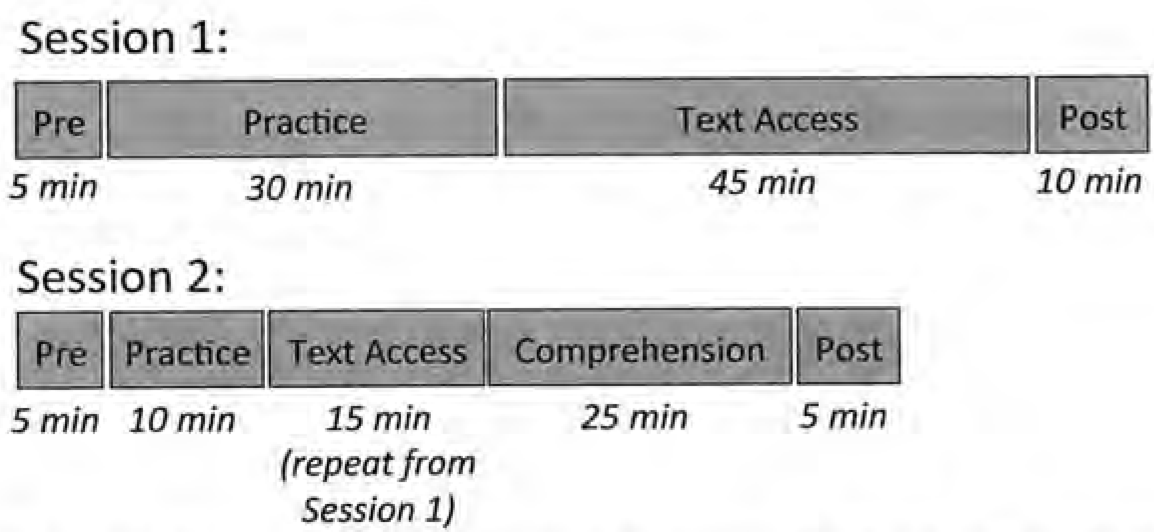


Figure 4-9: Overview of the schedule for both sessions. "Pre" and "Post" refer to interviews held before and after the tasks.

4.5.3.3 Results

The collected data was analyzed with an open coding approach by both experimenters independently. Please note that the blind experimenter coded the audio track of the recordings. In the following, we report on the findings with regards to our two goals set above, as well as general findings from observations, interviews and questionnaires.

Usage analysis: We observed a variety of interaction strategies that the participants employed to both access and read text. We broadly distinguish between two phases: *calibration* and *reading*. We also report on other strategies we encountered throughout the sessions.

Calibration phase: We observed participants perform different interactions, depending on the text structure, to determine the orientation and to "zero in" on the document.

In case of a single column document (like the training text and the news articles), we observed a "framing" technique: all participants first examined the physical document to estimate its boundaries, then indicated the top left border of the document with their non-dominant hand and gradually moved the FingerReader downwards from there until they found the first word. They then placed the non-dominant hand to that very spot as an aid to indicate and recall the beginning of the line.

In documents with a complex layout (like the business cards and the restaurant menu), all participants employed a "sweeping" technique: they swept the FingerReader across the page, wildly, to see whether there is any feedback, i.e. any text is being read back to them. As soon as they discovered text, they again placed the non-dominant hand at that spot to indicate the start of the text. Sweeping was also used in case the participant was not interested in the particular text passage he was reading to find a different passage that might draw his interest (e.g. to move from entrees to desserts on the restaurant menu).

Reading phase: After the calibration phase, participants started to read text at the identified position. All of the participants then traced the line until the end of line cue appeared. With their non-dominant hand still indicating the beginning of the line, they moved the FingerReader back to that position and then moved further down until the next word was being read back to them. When the next line was clearly identified, the non-dominant hand was again placed at the position of the new line. We observed all participants skip lines, particularly on the restaurant menu, forcing them to backtrack by moving upwards again. However, P1 had much less trouble interacting with complex visual layouts than P2 and P3, resulting in only little line skips.

P2 and P3 also employed a "re-reading" technique, moving the FingerReader back and forth within a line, in case they could not understand the synthesized voice or simply wanted to listen to a text snippet again.

Other strategies: P2 had issues with maintaining a straight line in session 1 and thus used another sheet of paper which he placed orthogonal on top of the paper to frame the straight line of the text. He then simply followed that line and could read quite well. He did not use any guiding techniques in session 2 because he wanted to experiment without that scaffold as it was "too much effort" (P2).

We also observed P1 using the FingerReader from afar, i.e. lifting the finger from the page and sweeping mid-air. He performed this technique to quickly detect whether there was text on the document, e.g. to see whether a business card was properly oriented or whether he was looking at the back of the card (i.e. no text being read back to him). As soon as he was certain that the FingerReader was picking up lines, he circled in and began with the calibration phase.

Observed exploration effectiveness: All participants found the business cards and newspaper articles easiest to access and read. All participants were able to read all of the business cards properly (i.e. names, affiliations/job titles and telephone numbers). They also managed to get the gist of 4 out of 5 newspaper articles (with each missed article being different per participant). The pamphlet was also perceived as easy to access with pictures being recognized as blank space.

All participants had difficulties in exploring the restaurant menu. Particularly P2 and P3 had issues with accessing the multi-column layout, therefore constantly using the sweeping technique to find accessible text. Comparing the observed performance across sessions, the experimenters observed that P1 and P3 showed an improvement in performance during session 2, only P2 was worse as he neglected to follow his guiding technique in the second session. The improvement was also underlined by comments from P1 and P3, as they found it easier to get used to the FingerReader in session 2.

Errors: The usage analysis also revealed a set of errors that occurred during text exploration with the FingerReader. As the amount of errors is not quantitatively representative, we choose to report on their quality. We subdivided errors into 4 categories: i) character misrecognized (Levenshtein distance < 2), ii) wrong word recognized, iii) entire word misrecognized, iv) false positive.

Qualitatively, errors in categories i) and ii) were never an issue, participants could make sense of the word as long as the context was recognized. Errors in category iii) led to the "re-reading" technique described above; the same holds for misrecognized consecutive words. The most severe errors were those in category iv). These typically occurred when the finger movement speed exceeded the fixed reading speed of the text-to-speech engine. Thus, the FingerReader was still busy uttering words when the participant had already reached the end of a line. To state an exemplary observation: In case of the restaurant menu, that contained only sparsely laid out text, they were reading "ghost text" in an area where there was no text at all. Consequently, revisiting that particular area at a later point provided no feedback and thus confused the user.

General findings: We report on general findings from the observations and interviews. Last, we report on the results from the post-study questionnaire from session 2.

- Visual layout: The restaurant menu and the business contained were typeset in different layouts, e.g. a multi-column one. This was particularly challenging for the participants, less so for P1. P2 and P3 were specifically challenged by the multi-column layouts, as e.g. "formattings do not exist" (P2).

	P1	**P2**	**P3**
General			
The overall experience was enjoyable	3	2	3
Accessing Text with FingerReader was easy	3	5	4
Reading with the FingerReader was enjoyable	3	1	2
Reading with the FingerReader was easy	2	1	2
Difficulty			
Accessing the menu was easy	2	1	2
Accessing the businesscards was easy	1	4	3
Accessing the newspaper articles was easy	4	1	3
Comparison to other mobile text reading aids			
Accessing text with the FingerReader felt easier			4
Reading with the FingerReader felt easier			3
Independence			
Felt greater desire to become able to read independently while on the move	2	4	3
Feel the desire to use the FingerReader to access text on the go	2	1	3

Table 4.3: Results from the questionnaire on 5-point Likert scale (1=stronlgy disagree, 5=strongly agree). The comparison to other mobile text reading aids was only applicable to P3.

- Synthesized voice: The voice from the employed text-to-speech engine was difficult to understand at times. Particularly P3 mentioned that it was hard for him to distinguish the voice from the audio feedback and thus missed a bunch of words occasionally. This lead him to employ the re-reading technique mentioned above.

- Audio feedback: P1 and P2 preferred the audio feedback over the tactile feedback and wished for an audio-only mode. P2 mentioned that the choice of audio feedback could be better, as he found it hard to distinguish high-pitch tones (line deviation) from low-pitch tones (finger twisting/rotation) which he called the "High-Low-Orchestra".

- Fatigue: All participants reported that they would not use the FingerReader for longer reading sessions such as books, as it is too tiring. In this case, they would simply prefer an audio book or a scanned PDF that is read back, e.g. using ABBYY FineReader (P1).

- Serendipity: Whenever any of the participants made the FingerReader read the very first correct word of the document, they smiled, laughed or showed other forms of excitement–every single time. P2 once said that is an "eye-opener". P1 said that it is "encouraging".

Table 4.3 shows the results for the post-study questionnaire from session 2. The overall experience with the FingerReader was rated as mediocre by all participants. They commented that this was mainly due to the synthesized voice being unpleasent and the steep learning curve in session 1, with session 2 being less difficult (cf. comments above).

The participants found it generally easy to access text with the FingerReader, while actual reading was considered less enjoyable and harder. All participants struggled accessing the menu. Accessing businesscards was easy for P2 and P3, while newspaper articles were easy to access for P1 and P3.

When comparing the FingerReader to other mobile text reading aids, P3 found that accessing text with the FingerReader was easier, yet he found reading with the FingerReader was comparable to his current text reading aids. He commented that he would use FingerReader for text exploration, while he would still want to rely on TextGrabber and Prizmo on the iPhone to read larger chunks of text.

Last, P2 and P3 felt a greater desire to read independently on the move, yet are torn whether they want to use the FingerReader. P2 and P3 commented on the latter that they would definitely use it in case they could customize the feedback modalities and have a more natural text-to-speech engine.

4.6 Discussion

In this section we discuss the results from the evaluation and highlight lessons learned from the development of the FingerReader. We hope that these insights will help other researchers in the field of finger-worn reading devices for the blind and inform the design of future devices.

Efficiency over independence: All participants mentioned that they want to read print fast (e.g. "to not let others wait, e.g. at a restaurant for them to make a choice", P3) and even "when that means to ask their friends or a waiter around" (P1). Though, they consider the FingerReader as a potential candidate to help them towards independence, since they want to explore on their own and do not want others suggest things and thus subjectively filter for them (e.g. suggesting things to eat what they think they might like). From our observations, we conclude that the FingerReader is an effective tool for exploration of printed text, yet it might not be the best choice for "fast reading" as the speed of the text synthesis is limited by how fast a user actually flows across the characters.

Exploration impacts efficiency: The former point underlines the potential of FingerReader-like devices for exploration of print, where efficiency is less of a requirement but getting access to it is. In other words, print exploration is only acceptable for documents where (1) efficiency does not matter, i.e. users have time to explore or (2) exploration leads to efficient text reading. The latter was the case with the business cards, as the content is very small and it is only required to pick up a few things, e.g. a particular number or a name. P2, for instance, read his employment card with the FingerReader after finishing the business cards task in session 1. He was excited, as he stated *"I never knew what was on there, now I know"*.

Visual layouts are disruptive: The visual layout of the restaurant menu was considered a barrier and disruption to the navigation by P2 and P3, but not by P1. All of the three participants called the process of interacting with the FingerReader "exploration" and clearly distinguished between the notion of *exploration* (seeing if text is there and picking up words) and *navigation* (i.e. reading a text continuously). Hence, navigation in the restaurant menu was considered a very tedious task by P2 and P3. Future approaches might leverage on this experience by implementing meta-recognition algorithms that provide users with layout information. A simple approach could be to shortly lift the finger above the document, allowing the finger-worn device to capture the document layout and provide meta-cues as the user navigates the document (e.g. audio cues like "left column" or "second column").

Feedback methods depend on user preference: We found that each participant had his own preference for feedback modalities and how they should be implemented. For instance P1 liked the current implementation and would use it as-is, while P2 would like a unified audio feedback for finger rotation and straying off the line to make it easily distinguishable and last, P3 preferred tactile feedback. Thus, future FingerReader-like designs need to take individual user preferences carefully into account as we hypothesize they drastically impact user experience and effectiveness.

Navigation during reading phase exposes the characteristics of navigation in an audio stream: The observed interaction strategies with the FingerReader indicate that navigating within text during the reading phase is comparable to the navigation in audio streams. The FingerReader recognizes words and reads them on a first-in, first-out principle at a fixed speed. Consequently, if the FingerReader detects a lot of words, it requires some time to read everything to the user.

This leads to two issues: (1) it creates noise, e.g. P1 and P2 frequently said *"hush, hush"* thus stopping the movement which interrupted their whole interaction process and (2) the mental model of the blind user–the respective cognitive map of the document–is specifically shaped through the text that is being read back.

As the speech is output at a fixed speed, the non-linear movement speed of the finger does not correlate with the speech output. Thus, any discrepancy between the position of the finger and the spoken text skews the mental model of the user. It is therefore important to establish a direct mapping between the interaction with the physical document and the speech output to maintain a coherent mental model of the document. This way, a direct interaction with the document would translate to a direct interaction with the speech audio stream. We suggest to employ adaptive playback speeds of the speech synthesis, correlating with the movement speed.

4.7 Limitations

The current design of the FingerReader has a number of technical limitations, albeit with ready solutions. The camera does not auto-focus, making it hard to adjust to different finger lengths. In addition, the current implementation requires the FingerReader to be tethered to a companion computation device, e.g. a small tablet computer.

The studies presented earlier exposed a number of matters to solve in the software. Continuous feedback is needed, even when there is nothing to report, as this strengthens the connection of finger movement to the "visual" mental model. Conversely, false realtime-feedback from an overloaded queue of words to utter caused an inverse effect on the mental model, rendering "ghost text". The speech engine itself was also reported to be less comprehensible compared to other TTSs featured in available products and the audio cues were also marked as problematic. These problems can be remedied by using a more pleasing sound and offering the user the possibility to customize the feedback modalities.

4.8 Conclusion

We designed FingerReader, a novel concept for text reading for the blind, utilizing a local-sequential scan that enables continuous feedback and non-linear text skimming. Motivated by focus group sessions with blind participants, our method proposes a solution to a limitation of most existing technologies: reading blocks of text at a time. Our system includes a text tracking algorithm that extracts words from a close-up camera view, integrated with a finger-wearable device. A technical accuracy analysis showed that the local-sequential scan algorithm works reliably. Two qualitative studies with blind participants revealed important insights for the emerging field of finger-worn reading aids.

First, our observations suggest that a local-sequential approach is beneficial for document exploration–but not as much for longer reading sessions, due to troublesome navigation in complex layouts and fatigue. Access to small bits of text, as found on business cards, pamphlets and even newspaper articles, was considered viable. Second, we observed a rich set of interaction strategies that shed light onto potential real-world usage of finger-worn reading aids. A particularly important insight is the direct correlation between the finger movement and the output of the synthesized speech: navigating within the text is closely coupled to navigating in the produced audio stream. Our findings suggest that a direct mapping could greatly improve interaction (e.g. easy "re-reading"), as well as scaffold the mental model of a text document effectively, avoiding "ghost text". Last, although our focus sessions on the feedback modalities concluded with an agreement for cross-modality, the thorough observation in the follow-up study showed that user preferences were highly diverse. Thus, we hypothesize that a *universal* finger-worn reading device that works uniformly across all users may not exist (sensu [211]) and that personalized feedback mechanisms are key to address needs of different blind users.

The next chapter will introduce the next step of evolution in our finger-worn assistive devices: the Mobile-FingerReader, a fully mobile version of the FingerReader that works as a smartphone peripheral. Learning from the shortcomings of the FingerReader, with the Mobile-FingerReader work we performed broader user studies with more VI persons, gathered quantitative data as well as redesigned the form factor.

Chapter 5

The Mobile-FingerReader

A Finger-Wearable Peripheral for Accessing Text with a Smartphone

Accessing printed text in uncontrolled environments such as outside the home is difficult for people with visual impairments. Existing smartphone-based technology presents problems of focus, aim and acquisition completeness. Continuing our work on small finger-worn devices that assist in accessing printed text, we set to achieve this goal in a truly mobile context. The new finger-worn device presented in this chapter connects to a standard mobile phone as a peripheral, and similarly provides online guiding cues for scanning the text. We report on the results of a user study to assess the feasibility of the Mobile-FingerReader, and the technical properties of the system.

5.1 Introduction

The Mobile-FingerReader presented in this chapter is a continuation of our work to enable access to printed reading material in an unstructured or unfamiliar environment in a mobile context for people with visual impairments (VI). The interviews with people with a VI in the previous chapter reveal that they struggle with focusing, aligning and even using assistive technology in settings such as in restaurants or reading mail items. These needs and problems were reiterated by interviewees who used the Mobile-FingerReader and appeal to the necessity for text-access technology that can overcome the hurdles of lighting, focus, aim and environment.

Extending the discussion from Chapter 4 we now focus on mobile efficacy via the usage of personal devices (e.g. smartphones). While it is reported to be important for people with a VI, in the past researchers suggested that mobile technologies present as much a challenge as a benefit [90]. In contrast to standard smartphones apps that offer text reading by use of the back-facing camera, our Mobile-FingerReader is a peripheral camera device worn on the user's finger to allow for finer control. The Mobile-FingerReader provides continuous real-time audio cues on scanning a line of text while speaking the words out loud to enable reading the print. Using a standard smartphone is key, since these devices are both prolific within the VI persons community and have ample computation power in recent generations. A peripheral device, which could be made cheaper as it uses less components, can spare the user from purchasing a costly specialized device or even a new smartphone by simply adding external capabilities. Peripheral and complementary devices are welcome in the VI community, a recent survey shows [260], as Bluetooth-coupled headsets and braille displays and keyboards are in wide use.

To evaluate our proposed text-reading smartphone peripheral we designed a usability study with 10 VI persons in a lab setting. Following our work from the last chapter

Figure 5-1: The Mobile-FingerReader camera peripheral

on the FingerReader, we sought to estimate the potential success of the device to aid in reading printed material. To supplement the former studies, we performed a quantitative assessment of the complete working system with a larger user base and the responses from interviews with the participants.

Our findings show that users were able to successfully extract an average of 74% of the words in a given piece of text when only provided with a feedback that told them how far away from the text line they were. The results demonstrate robustness in handling a range of standard font sizes, and that reading text within this range does not significantly hinder reading capability. The data also reveals insignificant advantage for residual eyesight when using the Mobile-FingerReader for reading, as some totally blind users actually had more success in reading than users with some residual vision.

The contributions of this chapter are: (1) a detailed recounting of the technical aspects of the Mobile-FingerReader, (2) results of a user study performed with 10 VI persons recording both quantitative and qualitative data and (3) a discussion around the implications of the Mobile-FingerReader as an assistive technology for accessing print.

5.2 Related Work

Much of the background on assistive finger worn camera devices was adequately covered in the Background chapter (Chapter 2) and the last chapter describing the FingerReader (Chapter 4); therefore in this section we will only briefly mention interesting work in the related fields.

Section 4.2.3 discussed many of the existing specialized products to assist in reading, however in the domain of smartphone applications there are the prolific LookTel[1] and Prizmo[2] among others. Crowdsourcing and peer-to-peer assistance via a mobile application is also on the rise in recent years providing help not only in reading but any visual task. The VizWiz platform [20] was the first widely used assistive application, however since then BeMyEyes[3] introduced assistance by video rather than an image, and TapTapSee[4] incorporated algorithmic recognition.

Beyond academic works that were already mentioned in Chapters 2 and 4 (see Table 4.1), some work not involving computer vision, such as El-Glaly's finger-reading iPad [42] and Yarrington's skimming algorithm [259], demonstrate the need to create an equilibrium between visual and non-visual readers by importing aspects of visual reading to assistive technology for VI persons.

[1]http://www.looktel.com/
[2]http://www.creaceed.com/prizmo
[3]http://www.bemyeyes.org/
[4]http://www.taptapseeapp.com/

5.3 The Mobile-FingerReader

The Mobile-FingerReader is a finger-worn peripheral camera device that is worn on the index finger. It works by recognizing any text seen right above the tip of the reader's finger. This is achieved through a pipeline of computer vision algorithms similar to the ones described in Chapter 4, but with modifications and improvements. We hereby detail the hardware and software aspects of the device, as well as the user interaction scheme.

5.3.1 Device hardware

Bearing resemblance to the FingerReader, the Mobile-FingerReader is designed to be smaller and better adjustable to differently shaped fingers. The 3D-printed plastic case sports adjustable rubber straps and ergonomic design for adhering to the top of the finger. It also contains a considerably smaller camera module than that of the FingerReader, although not as small as the HandSight's NanEye [226]. The Mobile-FingerReader, in contrast to FingerReader and HandSight, does not contain any vibration feedback capabilities and relies on audio cues alone, which allows it to be smaller and monolithic.

The camera module in use is analog; therefore a USB Video Class (UVC) video encoder is included with the system. The UVC interface allows the Mobile-FingerReader to connect to practically any device with USB host capabilities and a modern operating system, smartphones included. This way the Mobile-FingerReader, while currently still a prototype, could be used in the future as a peripheral by anyone carrying a smart device, e.g. a phone or an Android-enabled CCTV magnifier.

5.3.2 Android application

Accompanying the Mobile-FingerReader hardware we contribute an implementation of the original computer vision algorithms in Section 4.4.2 for the Android platform. The application, pictured in Figure 5-3b, serves as the main interface through which the external Mobile-FingerReader hardware can be controlled and the source of computation for processing the incoming video frames.

Through the application, a variety of settings are available to the user that enables him or her to customize their reading experience. Feedback settings can be adjusted to the user's preference, providing for the enabling and disabling of candidate line, distance, and angular feedbacks, as well as customizing whether incoming words are read in their entirety or cut off when a new word is found. Speech rate, the speed at which words are read, can also be adjusted.

For the purposes of aiding the user, all meaningful actions can be controlled through directional swipes on the main screen. Swiping upwards turns on the attached Mobile-FingerReader camera and begins the processing of video data and feedback. Swiping downwards stops this process and turns the camera off. Swiping right takes users to the preference menu. All other visible controls are presented for the purposes of aiding in debugging for testing purposes and present no further use to the user.

5.3.3 Computer vision algorithm

The bulk of the algorithms used for the Mobile-FingerReader are the ones used in the FingerReader, however our work contains a number of additional features and improvements. In this section we only highlight the novelty in our own work, and give a general overview of the system. In broad strokes, the reading algorithm is a state machine that governs the process of giving feedback to the user with the following states: *No Text*, *Candidate Line*, *Reading Line*, *End of Line* and *Line Lost*.

Existence of text (*No Text / Candidate Line* states) is determined by the number of qualifying character contours in the focus region, which is determined by the visible tip of the user's finger in the camera frame. If there are more than 2 qualifying characters that form a mutual baseline (tested by means of voting and fitting a line equation) the system transitions to *Candidate Line* state. In *Candidate Line* mode it will look for the first word on the candidate line via OCR.

The OCR engine, based on Tesseract [220], compensates for the distortion caused by the angle the finger takes with the paper. If the text is at an angle w.r.t the image, determined by the precomputed line equation, a 2D central rotation will correct it. Thereafter an intelligent trimming process will remove the whitespace surrounding the first word. We determine the first word by looking for large gaps in the x-axis projection of the words image patch (reducing the rectangular patch to a single row with the MAX operator on each column), similar to [226]. The trimmed patch is small enough to be quickly processed by Tesseract when set to the *Single Word* mode. OCR also does not occur on every frame, rather, only when new candidate words appear, greatly improving performance on our mobile processor.

The finger-tip detection algorithm of the FingerReader was inefficient and expensive to execute in a mobile setting. We therefore introduce a coarse-to-fine method, where we start by analyzing an extremely downscaled (uniform 10%) version of the normalized-RGB image and later inspect the rough estimate in a small 100x60 pixel window to get a precise reading. We also incorporate a standard Kalman filter to cope with noise in the measured fingertip point signal, which has a detrimental effect on the stability of the algorithms down the pipeline.

	Age	Gender	Hand	Visual Impairment	Print access
P0	61	Female	Right	Retinopathy of prematurity. Congenital. Totally blind	PC: Flatbed scanner, Kurzweil Mobile: kNFB reader, Prizmo, SayText
P1	20	Male	Right	Extreme myopia (¡ -12.00D)	Uses heavily corrective glasses, reading very close
P2	30	Female	Right	Retinal damage. Congenital. Able to see shadows	PC: OpenBook Mobile: kNFB reader, StandScan
P3	63	Male	Right	Cancer of the retina. Totally blind.	PC: TravelScan, ScanSnap Mobile: kNFB reader, braille transcriptions.
P4	36	Female	Right	Stevens-Johnson syndrome. Onset at age 3. Some light perception.	PC: Flatbed scanner Mobile: kNFB reader, human readers.
P5	61	Male	Right	Retinal cancer. Blindness onset at age 9. Totally blind.	PC: Flatbed scanner Mobile: kNFB reader, AbiSee scanner, Kindle, Optacon.
P6	63	Female	Right	Retinopathy of prematurity. Congenital. Totally blind.	PC: OpenBook Mobile: kNFB reader, AbiSee scanner, Kindle, BookShare.
P7	54	Female	Right	Macular scatoma. No central vision.	PC: ScreenReader Mobile: CCTV magnifier, magnifying glass
P8	39	Male	Right	Retinopathy of prematurity. 20/200 right eye, left eye a prosthesis.	PC: ScanSnap Mobile: kNFB Reader
P9	34	Male	Left	Retinitis pigmentosa. Totally blind.	PC: Flatbed scanner. Mobile: kNFB Reader, human readers

Table 5.1: Details of the participants in our user study. Note: P0 was a pilot participant.

5.3.4 The feedback mechanism

Each working state emits a different audio feedback to guide the user: *No Text* and *Candidate Line* emit a continuous square wave with different pitch, the *Reading Line* state emits a changing continuous sine wave, which is distinguishable in timbre from the non-reading modes. The *End of Line* emits a looping "Beep" sound and *Line Lost* emits a singular "Beep".

To signal going above or below the line of text, once such a line was found, the smartphone emits a continuous sine wave of gradually rising/falling pitch: higher pitch for high position, lower pitch for low position. We determine the distance of the fingertip from the line in image-pixels using a simple point-to-line distance calculation. When the fingertip is close to the line this feedback is reduced in volume. We also similarly signal the angle the finger creates with the line of text, using a higher range of pitch for that sine wave.

5.4 Evaluation

Investigation of the FingerReader did not include quantitative measurements (see Section 4.5), and other similar systems did not evaluate an end-to-end system [226]. The primary contribution in this chapter is a report of a quantitative assessment of the complete system, including computer vision subsystem, as used by a larger group of visually impaired persons.

5.4.1 Runtime Analysis

The Android-based implementation of the text reading algorithms reached an average frame rate of 5-10 frames per second. A runtime analysis, which can be seen in Figure 5-2, shows that tracking the words through the scene (described in Section 4.4.2.1) adds considerable computation load that scales linearly. Tracking the line however did not present a significant change from only tracking the finger.

5.4.2 Evaluating With Users

To evaluate the usefulness of the feedback mechanisms and overall usability of the system, we recruited 10 participants to undertake monitored usage tasks and interviewed them about their experience. In total 10 tasks were designed to contain text of different sized fonts (6pt to 12pt), layout (centered, left and right aligned, justified) and two variations of the audio feedback.

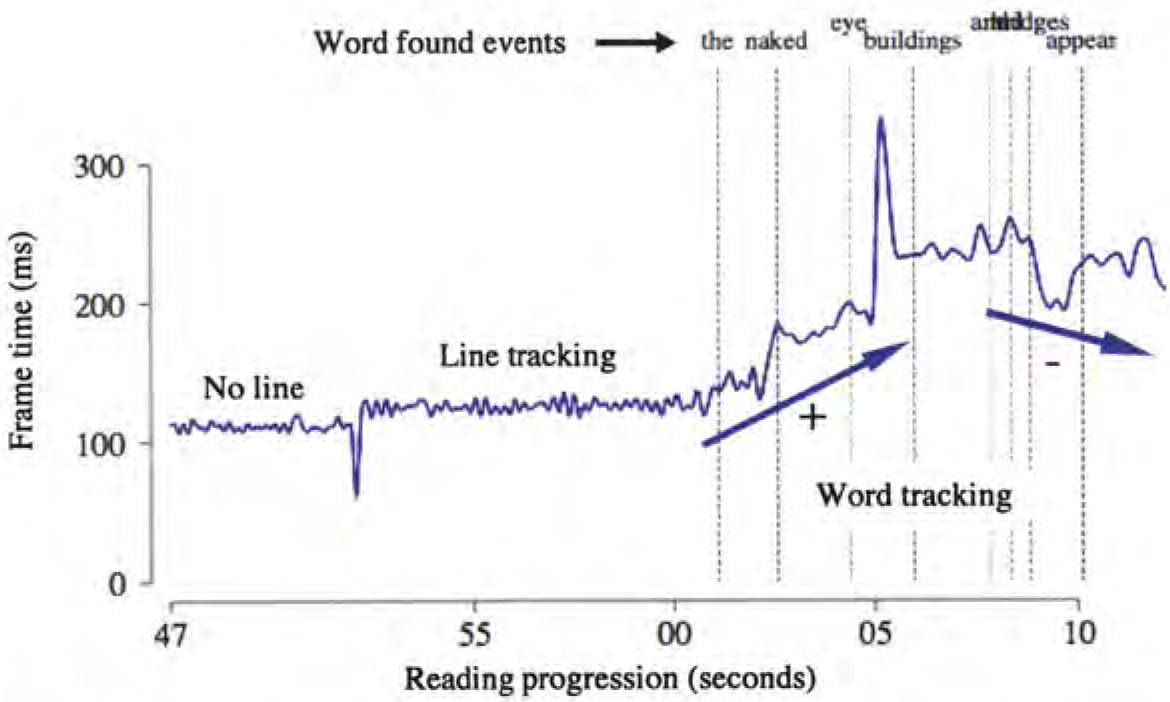


Figure 5-2: A frame-rate analysis of the reading progression, taken from a reading session of a VI study participant. As more word are tracked in the scene the total time for computation rises, and as they phase out of the scene it reduces.

(a) (b)

Figure 5-3: (a) P7 in midst reading with the Mobile-FingerReader, (b) Android app screen

5.4.3 Participants

Nine participants (4 female, 5 male, aging 46 ± 15 years on average, range 20-63) were recruited from a ready pool of volunteers. Details of the nature of visual impairment and their text-reading habits are given in Table 5.1. An additional participant volunteered to be a pilot participant (in Table 5.1 appears as P0) and helped reduce the bugs and oddities in the system, as well as practice the study procedure. P9 refused to perform many of the study tasks since he did not see the benefit of using the Mobile-FingerReader for reading, however he did spend enough time using the device to warrant his inclusion in the qualitative feedback and interview. All participants received compensation for their time.

5.4.4 Procedure

The study contained four major parts: Pre-usage questionnaire, Practice, Usage tasks, and a Post-usage interview. A typical duration for a single participant was 90 minutes.

The practice session included a demonstration of the device and the audio feedbacks. A printed text sheet with a single line of text in one part and two lines of text in another part, was used to practice staying on one line and finding a subsequent line. During the practice the participants received full support in finding the text and working with the feedbacks to stay on it. Thereafter the participants were given 10 reading tasks consisting of 8 news abstracts (averaging 4.5 lines and 48.4 words with a SD=4.5, ranging the font sizes: 9pt, 10pt, 11pt and 12pt) and 2 mock-business-card (each 8 lines and 22, 23 words, font size 6pt to 12pt). Participants were given up to 5 minutes to complete a single task, and did not receive any assistance from the investigators save for encouragement.

We tested two feedback conditions with each subject: Distance (D) and Distance+Angle (D+A). In 'Distance' the user hears a continuous feedback of how far their fingertip is from the line, and in 'Distance+Angle' the users also hears a continuous feedback of the angle their finger makes with the line. Both feedbacks were given as sine waves of different pitches (Distance: 540Hz to 740Hz, Angle: 940Hz to 1140Hz). Each feedback condition was crossed with the tasks (5 tasks for D and 5 tasks for D+A) and fully counterbalanced to remove order bias.

To gather qualitative feedback we performed semi-structured interviews starting with a Likert scale questionnaire (15 questions), 3 open questions and finishing with an open-ended discussion about the experience and beyond it.

5.4.5 Setup

We replicated the study setup of [226] with the iPad and overlaying paper; however our study was done using the real wearable computer-vision system instead of a simulation driven by the iPad's touchscreen. We programmed a Node.JS web application to use on the iPad to track finger position. The reading task paper was mounted and aligned

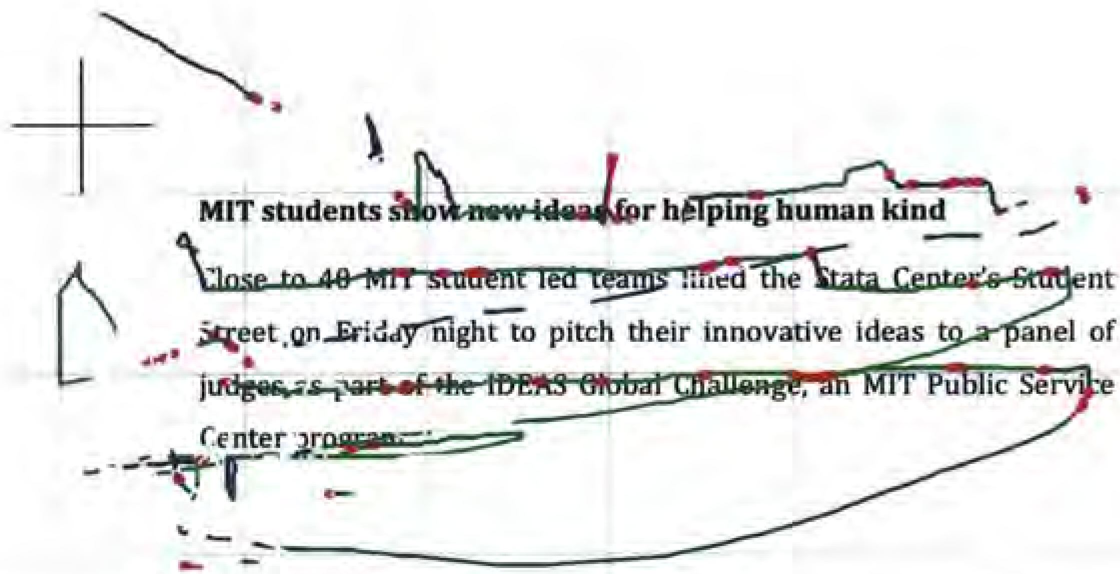


Figure 5-4: Aligned data from the iPad of a single session for P4, a totally blind person. The green-blue marking are a visualization of the *Distance* feedback the user received, while the red marking is when word were spoken out loud. Greener hues correspond to a closer distance to the baseline, and blue hues farther. The middle line of the text was missed, however the user was still able to recover 73% of the words in total and spend over 51% of the time line-tracking.

on the screen using a printed anchor point so the printed text can later be aligned with the recorded touchscreen data (see Figure 5-4). Multiple parts of the user's hand could be touching the paper and sensed by the iPad. Thus, the application only tracks the top touch point from the user, corresponding to the touch of the finger wearing the reading device. The application records the x-position, y-position, and time-stamp of each finger movement over the paper and writes that information to a text file. It also draws the finger's path on the iPad screen for easy, immediate visualization. The reading device was attached to a Samsung Galaxy S5 Android phone placed on the table, which was emitting the audio feedbacks (see Figure 5-3a).

5.4.6 Analysis

The quantitative measurements were aligned, cleaned and analyzed as detailed in the following sections. The qualitative feedback was transcribed, coded and categorized by the investigators during the interviews and later from the video recordings.

5.4.6.1 Data synchronization and alignment

Since our data was recorded simultaneously on two devices (Smartphone and iPad) we needed to synchronize the data streams. Millisecond timestamps were recorded for each stream event, and later lined up in single time-series following the events from the phone (state changes and audio feedback to the user) by interpolating the numerical tracking data from the iPad (x-y position of the finger).

To align the 2D positioning from the iPad with the printed text, we used the known resolution of the iPad (132 DPI) and the on-screen offset for the printed mark in pixels. First offsetting then normalizing for the resolution, we align programmatically with a rendered PDF of the page in 200 DPI. The results of one such alignment can be seen in Figure 5-4. For purpose of visualization alone in Figure 5-4 we added an offset to move the touch data to coincide with the text, where the original position was much lower (that accounts for the camera's center of projection and the difference between the fingertip and pad of the finger where the touch happens).

5.4.6.2 Extracted measurements

From the collected and aligned raw data we extracted the following measurements:

- Reading proficiency (*"Consecutive Score"*): This measures the amount of correctly and consecutively extracted words from a piece of text compared to the ground truth (list of known words in the text). The final score is based on a histogram on the length of correct words sequences. For example, if P1 was able to read 3 correct words in a row, the histogram value for a sequence of 3 gets increased by 1, and so on for any length of correct and consecutive words sequence. The final score for a single reading task is calculated as: $S = \sum_k H(k) * k^2$ with $H(k)$ being the histogram bin value for sequences of length k. This metric gives higher marks for longer sequences read, assuming long sequences create a better understanding of the text. We also discount for slightly misspelled words by allowing for an edit distance of 2 in the match (for words longer than 4 characters), assuming such slight misspellings do not inhibit understanding by much. We normalize the histogram-based score by the highest achievable score on a given task: reading it all the way through consecutively line-by-line.

- Word extraction proficiency (*"Total Words Read"*): This counts the number of words the user was able to extract from the print in relation to the amount of words in it, without regard for order. This gives us another perspective to text comprehension from the Mobile-FingerReader output, since the *Consecutive Score* measure allows for repeating sequences while this measure is robust to them.

- Following the feedbacks: This measures the amount of time participants spent in the tracking modes (*Reading Line* and *End of Line*) in relation to the time spent in *No Line* or *Candidate Line* modes. To calculate, we simply extracted the amount of time a user spent in the tracking modes and divided by the cap time for performing the task.

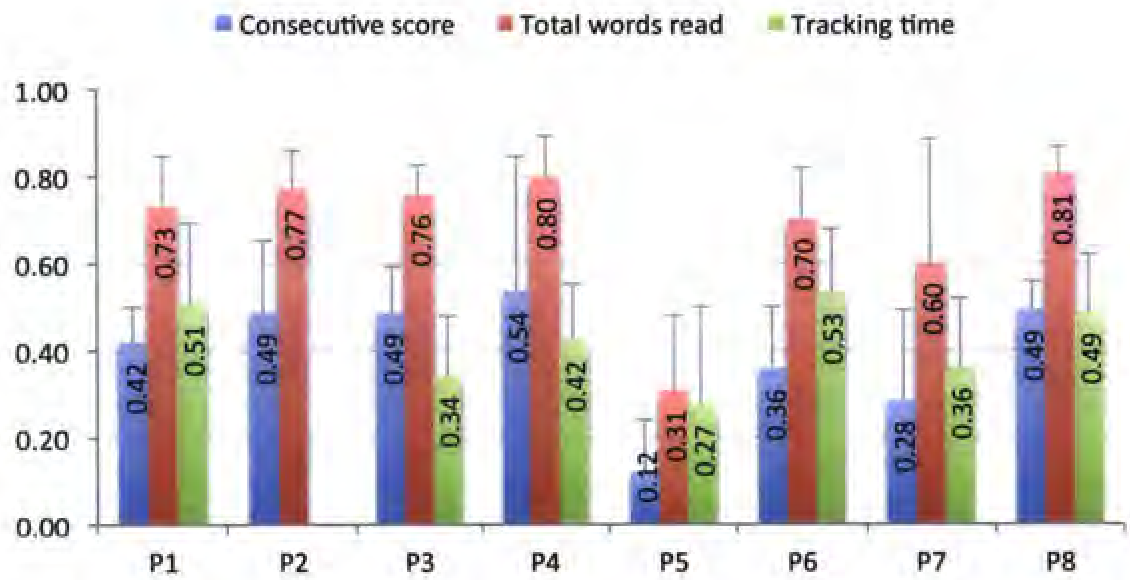


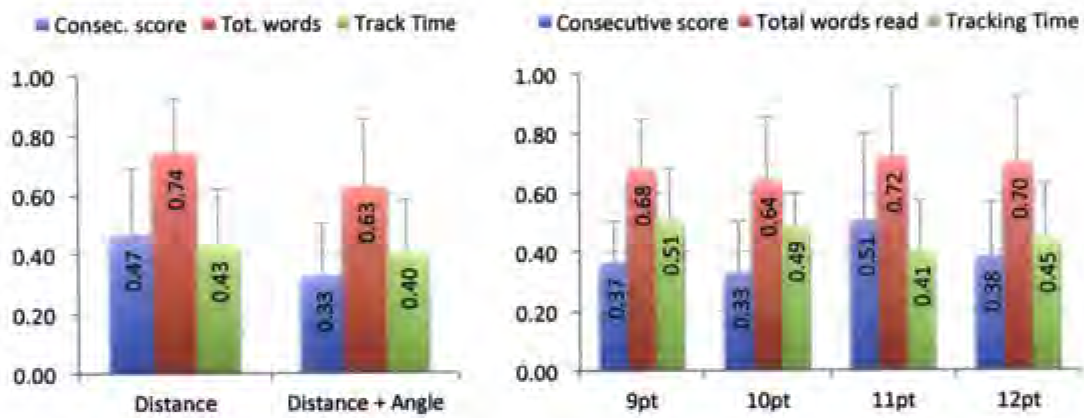
Figure 5-5: Individual success in reading per participant. Note: we were unable to extract timing data for P2.

5.5 Results

5.5.1 Quantitative measurements

Our results show that on average our participants were able to correctly extract 68% (SD=21%) of the words in the text, however some participants were able to extract up to 81% on average (see Figure 5-5). As expected, the users with residual vision had more success extracting words (e.g. P8, P1), however some totally blind users were also rather successful (e.g. P4, P2).

The Distance feedback was somewhat better in helping users extract words from the text with 74% (SD=18%) of the words on average, relative to 63% (SD=22%) for Distance + Angle ($p = 0.009$ in a one-way ANOVA). Bigger font size only had a small positive effect w.r.t percent of extracting words (e.g. 72% for 11pt and 68%



(a) Feedback type

(b) Font size

Figure 5-6: Reading success as crossed by (a) feedback type (a), and (b) font size.

for 9pt), but made a bigger impact in terms of the *Consecutive Score* (with 0.51 for 11pt and 0.37 for 9pt), which suggests, as one would expect, that larger font is easier to track. A one-way ANOVA on the "Total Words Read" measures resulted in a p-value of 0.116, which suggests font size did not have a strong effect on successfully extracting more words.

As the *Consecutive Score* is not an absolute measurement, but rather a suggested model of the proficiency of a user in utilizing the Mobile-FingerReader, it only can serve as a comparative measurement. As such, it does flush out the variance in users capabilities when it comes to feedback. Users not only extracted more words with only Distance feedback turned on, they were also capable of extracting more consecutive words, with a score of 0.47 vs. just 0.33 for Distance + Angle.

The *Tracking time* measure provided only little information as to how successful users were in utilizing the feedbacks given for tracking. Interesting to note that P6 was the best in terms of time spent in the tracking modes (53% of the time), probably due to the fact that he was an Optacon user. Additional analysis of these results exists in the following discussion section, Section 5.6.

5.5.2 Qualitative feedback

The qualitative feedback we collected after participants concluded the tasks was divided to parts: (1) Overall perceived success and enjoyment, (2) Perceived understanding of the audio feedback, and (3) Perceived independence.

While the experience of reading with the Mobile-FingerReader was perceived to be rather enjoyable (2.00 SD=1.00), it was not thought to be easier than other reading aids (3.90 SD=1.13). It was also not perceived to be very easy (3.45 SD=0.85), even though on average participants reported they could understand the text (2.30 SD=0.78).

The best perceived audio feedback was *End of Line* (1.50 SD=0.92), and Distance from the line was also well detected (2.10 SD=0.94). However the Distance + Angle combined audio feedback was considered to be confusing (3.75 SD=0.68).

In terms of independence, the participants generally felt that a technical person would not be needed to operate the Mobile-FingerReader (3.55 SD=1.31), and it wouldn't require a great deal of things to learn (3.40 SD=1.2). Some participants were impressed by the ease of use and the small form factor:

> P6: "I was [reading] with very little equipment (just a ring) and you only need to keep control of the finger. It gave instant feedback if you got the word right."

Open ended interviews with our participants revealed that all, save for one, did not appreciate the Angle feedback and were confused by multiplexing Distance + Angle (N=8); this is supported by their rating of this feedback and quantitative accuracy scores too (see Figure 5-6a). Most users (N=5) also mentioned the usage of the device causes excessive arm strain in keeping the finger and wrist straight and tense, as well

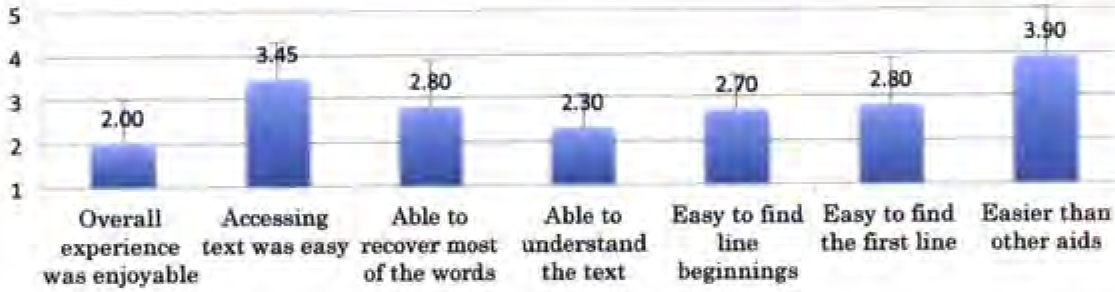


Figure 5-7: Qualitative measurement of perceived success. 1 - best, 5 - worst.

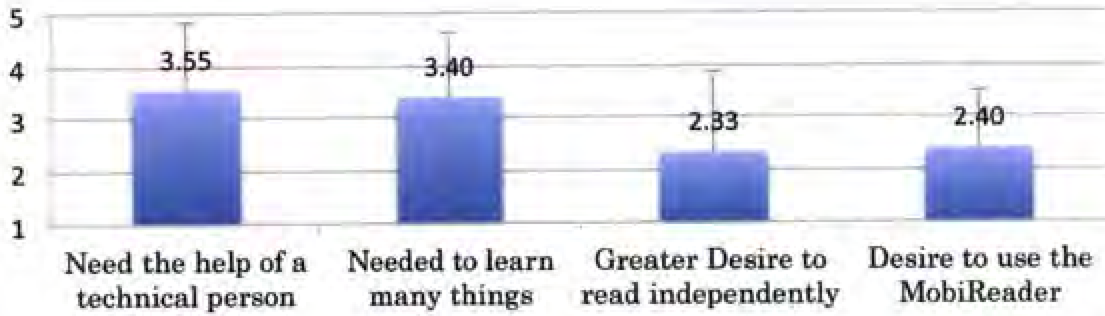


Figure 5-8: Qualitative measurement of independence. 1 - best, 5 - worst.

as having to be very accurate and make very slight constrained movements (N=3). Three users stated they would not use Mobile-FingerReader to read long pieces of text, even though it was generally agreed that the device design was comfortable and small (N=5). Some complained the overall reading process was slow (N=3).

The prevailing reported strategy (N=5) was to go top-to-bottom, i.e. finding the top line from the top of the page and working down to the next lines, as well as tracing backwards to the left to find the first word on the line; however backtracking was contested by some (N=3).

Some users expressed dislike for the feedback in general, claiming the tone and increasing volume when straying from the line induced more panic than suggestion. At times this was reflected by large movements that could throw the user off the current line.

Figure 5-9: Qualitative measurement of understanding the audio feedback. 1 - best, 5 - worst.

5.6 Discussion

While great strides are made to create finger-wearable devices for assisting people with visual impairments in reading and accessing visual material, the current state of the field, Mobile-FingerReader included, is that they still provide low accuracy and cumbersome operation preventing wide usage. Though our improvements in the Mobile-FingerReader have increased its usability, all our participants wished for a more accurate and smooth reading experience. P9 even went to ask us *"Why are you doing this?"* as he demonstrated that using the kNFB reader app he is far more efficient. This goes in accordance with our observations that when using a finger, VI users expect an immediate and lag-less feedback, as they would from braille. We therefore conclude that faster algorithms and a better user experience is still a pressing need.

The following is a discussion around the major elements that had impact on the reading success rates:

- **Skipped lines**: As Figure 5-4 shows, successful tracking of one line does not mean successful continuation to the next line. In transfer between lines, some users would over-shoot the next line, skipping to the line below and missing all the words on the first line. This was caused mainly due to lag in the *Candidate Line* signal, which the users did not perceive in time to stop and backtrack to the beginning of the line.

- **Misaligned signals**: Our runtime analysis (see Section 5.4.1) shows an effective response rates of 5-10Hz, which was too low to support fast reading motion of more than 2 words/sec. Many of our participants were moving their finger too fast for the system to process and provide guiding feedback. These situations resulted in confusion cases where the system was giving feedback for a scene it saw 100ms ago but the user's finger was already further away (sometimes too far above or below the line, or too late for adjusting). As noted before, overlapping tonal feedback for Distance+Angle introduced additional confusion in understanding and separating the signals.

  P6: "Sometimes you don't move and you can't tell why suddenly you're not hearing anything, even through I heard a word in the same area."

- **Environmental effects**: In our study setup we used proper bright lighting to assist the on-device camera in getting a well-lit scene of the page (as can be seen in Figure 5-3a). While in general the lighting had a positive effect, sometimes the hard shadows form an object, such as the user's other hand, fingers or wires from the device, eclipsing the light source would impede the computer vision pipe line. In addition, the positioning of the device on the users finger would occasionally drift (due to pulling on the wires coming out of the back) and misalign the camera with the finger, contributing to more errors in the scene analysis.

Our users expressed agreement that using the finger contains potential as a means or reading printed text, drawing connections to braille as an experiential parallel. Reflecting on our results, in the following section we propose ideas for how to improve the current state of the Mobile-FingerReader.

5.6.1 Future Work

The goals of this study focused primarily on investigating the viability of each audio feedback and assessing the overall reading experience for VI people. Future studies could explore user interaction with the main interface on the phone, focusing on the application usability. There is also room for investigation into the the robustness of the Mobile-FingerReader when applied in more mobile contexts.

Current Mobile-FingerReader hardware requires tethering to a power source for the ring. A future improvement that could remove this restriction and aid in the aforementioned future studies would be to use the phone's power source to power the mounted camera. UVC cameras powered strictly by the power source from the USB host are in abundance, however a very small module than can fit in a ring device must be developed.

Taking into consideration user inhibition of the angular feedback, it may be worth considering eliminating the feedback altogether and improving the automatic correction by the software component. Doing so would reduce the amount of cognitive effort necessary for a user to track a line, at the cost of performance on the application end. To put the usability of the Mobile-FingerReader on par with other market products, this may be necessary.

It is worth looking into adjusting the tones and frequencies emitted as feedback to prevent alarming users. P7 noted recent studies involving silent, electric cars by the U.S Department of Transportation have explored the domain of alerting sounds for the purposes of aiding visually impaired individuals in detecting approaching vehicles. For the purposes of improving the audible feedback of the Mobile-FingerReader, such studies may provide promise for providing phlegmatic audio feedback.

Users P5 and P6, both Optacon users in the past, suggested the development of a more constricting physical mechanism to facilitate line tracking with the hand, similar to the wheels on the Optacon that constrain it to lateral movement. They did admit that the Optacon took many months of training to master compared to less than an hour with the Mobile-FingerReader. As other users also struggled with keeping the hand in a correct pose for tracking, we believe a method to allow for a relaxed and casual reading is in order.

5.7 Conclusion

We presented the Mobile-FingerReader, a smartphone camera peripheral for reading printed material, and an evaluation of its usability with visually impaired persons. The Mobile-FingerReader presents a number of technical improvements over the FingerReader, and showed a greater potential to become a useful assistive device. Quantitative data from a study with 10 users on reading with such a device provided insight into the feedback modality and font size constraints. However even with the advancements made, we believe further research into the user experience in terms of control, algorithm and feedback is required in order to turn finger-worn device into a viable solution for reading print.

The next chapter will discuss a different application for finger-worn assistive devices - reading musical notation sheets. The chapter presents the technical details and user study work to evaluate the proposed music reading system, and also provides insights into the needs and wishes of VI musicians as well as the gaps in existing technological offering.

Chapter 6

The MusicReader

Assistive Optical Music Recognition from the Finger's Perspective

Paper musical notation sheets are considered highly inaccessible for blind music readers. They rely on stationary scanning equipment and aid from sighted users, somewhat impractical accessible solutions such as Music Braille or learning pieces by ear. This chapter proposes a new method for interactively reading printed musical notation sheets using a finger-mounted camera. Our real-time local-sequential stave notation recognition and tracking algorithm allows for a non-visual scanning of a musical sheet with the finger by receiving continuous audio feedback. The system was tested with blind musicians and shown to have potential in accessing a musical piece that would otherwise require a lengthy and cumbersome effort.

6.1 Introduction

Optical music recognition (OMR) is a booming scholarly endeavor of the last five decades that recently presented real-time and even mobile stave notation recognition systems. This stands in stark contrast to the inaccessibly of stave notation sheets for readers with visual impairments (VI). From interviews with music readers with a VI we learned that to be able to read printed music sheets they rely on human transcribers or scanning using specialized stationary equipment, which also often produces recognition errors. For students of music that participate in classroom or band sessions this issue creates a barrier between them and their sighted colleagues, as they are not as independent as their peers. Our proposed system, presented in this chapter, strives to enable VI music readers to access non-instrumented paper musical notation sheets in a mobile context, and level the playing environment with their sighted peers.

6.1.1 Needs Of Music Readers With Low Vision

Current solutions for VI music readers are either accessible formats of music notation or digitization of music sheets via scanning with specialized audio interface. Music-Braille is a relatively prolific accessible format for encoding musical information based on the Braille character set, however it presents a number of acute problems. Music-Braille is expensive to produce and thus also to purchase, since it is a niche format for musical notation, which leads to a small offering of music translated to Music-Braille [79]. Learning how to read Music-Braille is also a challenge as not so many teaching institutions exist (our interviewees know of only about 50 teachers in the entire U.S). Music-Braille translation also results in very heavy and large “printed” books, which is another usability factor impeding accessibility.

In light of these and other hurdles, VI music readers opt to learn music by ear or digitize printed music sheets using OMR, however that is also not free of limitations. Operating a flatbed scanner requires experience using a screen-reader and the specific print digitization software (for example SharpEye[1] or SmartScore[2]). The physical setup for scanning is also important for a successful sightless operation, therefore it is usually done in a recognizable comfortable location, such as a specialized room or at home. But even in perfect scanning conditions, a properly scanned page will often result in errors in the OMR process. Scanning in a different scenario, such as using a mobile phone, presents problems of aim, focus and alignment, but more importantly - such mobile music scanning applications for the VI are hardly in existence.

These problems hinder VI musicians from successfully scanning music in a mobile situation, and more importantly - in a classroom or a band, where they are expected to access music handouts in the same way their sighted peers do. Accessible workbooks for classroom usage do exist and pre-digitized music is also abundant online, however then VI readers are confined to this content and cannot spontaneously access other printed material that is otherwise not digitized. Consequently, a music reading solution tailored for the VI to use in a mobile context could provide them with a way to better integrate into the learning and communal playing environment.

6.1.2 Reading Music With The Finger

This chapter presents our efforts to create a mobile and easily operated music scanning device for VI music readers. Our current implementation is based on our previous work on finger-wearable devices for the VI that assist in reading printed text (see Chapters 3, 4 and 5). The index-finger wearable device contains a small camera that looks down the finger and analyses the print above the fingertip. This setup allows

[1] `http://www.visiv.co.uk/`
[2] `http://www.musitek.com/`

the reader to move the finger along the page without restriction, to skim and jump to different areas, in contrast to existing OMR solutions that dictate scanning a whole segment at a time. Our system also maintains the fingertip's touch of the paper to keep the natural tactile response of the surface of the paper, which is vital for non-sighted operation. This tactile feedback is lost when using a mobile phone to take a picture from afar, or when a special hand-steerable device (e.g. [10]) is touching the paper. Additionally, we provide continuous audio feedback on the scanning to guide the VI user in a non-visual manner. Real time audio feedback, of the kind our proposed system offers, is crucial in the creation of a proper mental model in the mind of a non-visual reader, as the layout and symbols on the page gradually construct a mental map.

We contribute the following unique elements to the growing field of mobile, real time OMR systems as well as assistive technology for VI musicians:

1. an algorithm for sequential extraction and tracking of musical information from a local camera viewpoint,
2. design of a finger-wearable device for scanning printed music,
3. continuous audio feedback tailored for VI music readers, and
4. evaluation of the system with the target audience that provides valuable insight into creating assistive music applications.

The following sections of the chapter present the background and technical details of these contributions.

6.2 Related Work

OMR is a steadily growing field of research of roughly the last 50 years that has reached a very advanced state with near perfect performance under certain conditions. For a review of the history of the field of OMR we refer the reader to [45], and for a thorough overview of the state-of-the-art and open problems in OMR the reader will find [177] an interesting read. In spite of years of research, OMR still presents a great wealth of problems to solve when: imaging conditions are not perfect, the view is partial or distorted, the print is irregular or hand-written, computation power is small, or in presence of other constraints such as mobile, real-time or accessible operation. Traditional aspects of OMR, such as staff line detection and removal, are still avidly researched, however the user applications arena for OMR is rapidly expanding.

In Table 6.1 we compare a number of attempts at OMR systems that provide real time or mobile operation, similar to our goals. The incredible feat of Matsushima et al. from 1984 in creating the WABOT-2 music-playing robot, exemplifies how OMR can be performed to high accuracy in near real-time speeds (authors claim a 100% accuracy on certain pieces, and 10 seconds processing time) [237]. More recently, a number of other approaches were suggested to deal with other needs. When the camera is unable or not intended to see the entire sheet, what we call "Semi-Local" and "Local" view (depending on the size of the visible part), the proposed systems tend to focus on extracting only notes to allow for playing them back immediately in real-time [10; 92] or within seconds [221]. Advancements in OMR also allowed for creation of mobile phone applications that perform the processing on the phone itself [243; 127] or with a fast network connection to a processing server [224].

Perhaps the closest piece of work to our wearable system is Gocen [10]. In Gocen the user is allowed to scan a stave notation line using a handheld camera and playback the

Work	Response	Playback	Device	View	Symbols
Matsushima [1985]	15s	Yes	Stationary (Robot)	Global	Many
Gakufu [2011]	Realtime	Yes	Mobile phone	Semi-Local	Notes
Gocen [2012]	Realtime	Yes	Handheld	Local	Notes*
Luangnapa [2012]	10-15s	No	Mobile phone	Global	Many
SnapNPlay [2013]	<10s	Yes	Mobile phone	Semi-Local	Notes
Soontornwutikul [2013]	10s	No	Mobile phone	Global	Many
iSeeNotes [2014]	<10s	Yes	Mobile phone	Global	Many
Our system	Realtime	No	Finger-wearable	Local	Many

Table 6.1: A comparison of existing real-time and mobile music reading solutions.
* The system strictly recognizes only full, stemless notes.

notes sequentially in real time. However in contrast to our approach, Gocen does not recognize any symbols other than full, stemless notes (essentially only note heads), has no memory of notes outside of its immediate focus region, and doesn't provide feedback on the scanning other than playing the note. Another interesting related work is onNote by Yamamoto et al. [255], where the user is allowed to use the finger to access different parts of a paper-printed music sheet as well as change the nature of playback. In onNote however, the pieces are scanned into the system beforehand and the only real time computer vision operation is that of feature-based matching to the existing database of pre-processed sheets. onNote uses a projector-camera system for visual feedback for the user, while we use a finger-wearable system and an audio feedback.

6.2.1 Assistive Solutions for Visually Impaired Musicians

While music braille [217], created by Louis Braille himself, and other forms such as large embossed stave notation exist as accessible format of music notation, they are not a complete solution for people with a visual impairment. In spite of many libraries offering downloadable music braille nowadays [79] and online services offer translation to braille [58], the problem of accessing music without a braille display or printout still remains. In Challis' broad body of work on music accessibility we found interesting propositions for an alternative to music braille, for example the

Weasel system [28] that uses both tactile and auditory cues. Crombie et al., who also contributed greatly to making music accessible to VI musicians, proposed Spoken Music [37] - a protocol to encode music in words, which joins the ABC[3] and Talking Scores[4] encoding mechanisms. Recently, Bajo et al. presented a vibrating bracelet to aid VI musicians as part of an orchestra [13] and Capozzi contributed Musica Parlatta as a method to vocalize the spoken music according to pitch (virtually "sing" Spoken Music) [26].

6.3 Finger Worn Reading System for Printed Music

Our music reading system consists of two major physical components: a small finger-wearable camera device, and a computation device. The wearable camera device is a continuation of our research into wearable assistive devices for the VI that focused on seamless natural access to printed or other visual information. The EyeRing (Chapter 3) provides access to price tags, currency notes and recognizing colors, while the FingerReader (Chapter 4) allows for accessing printed text while receiving continuous audio feedback. The finger device presented in this chapter is newly designed, and features an adjustable rubber strap for more comfortable wearing and a considerably smaller camera module (see Figure 6-1). The computation device we used is a laptop computer, however a mobile phone version of this system is in development at the moment.

The computational elements in our system are: a stack of computer vision algorithms, and a user interface designed for the VI. The computer vision algorithms, detailed in

[3] http://abc.sourceforge.net/

[4] http://www.rnib.org.uk/information-everyday-living-home-and-leisure-music-reading-music-accessible-formats/talking-scores

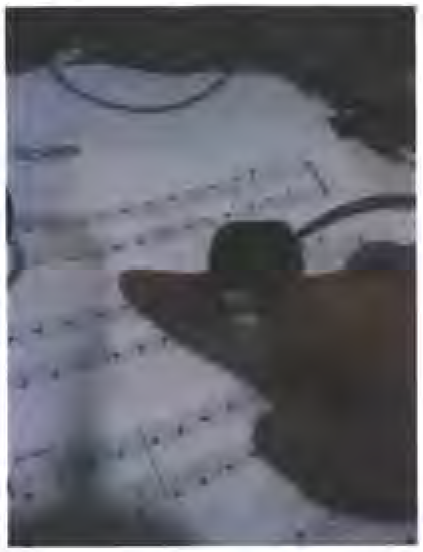


Figure 6-1: The wearable camera design, similar to the one presented in Figure 5-1.

the next section, receive the video feed from the finger wearable camera and analyze it in real time to detect musical symbols such as: staff lines, notes and their duration, time signature, accidentals, rests and clefs. To support a truly real time function, and create a strong perceptual link between the position of the finger and the system's continuous feedback, the algorithm considers only a small interest region in the image, derived from a detection of the fingertip. Music symbols in this small region are detected and reported to the user, in forms of speech and an audible tone. As symbols leave this small region of interest, by the motion of the finger across the page, they are tracked in order to prevent from considering them as new symbols and reporting them again to the user. This simulates the natural continuous sequential reading of the musical sheet in a collaboration between the user's finger and the computer's guidance.

6.3.1 Music Extraction From Finger Perspective Imaging

Our computer vision system considers a unique approach for extracting musical information from printed sheets. A finger-worn camera provides a local view of the

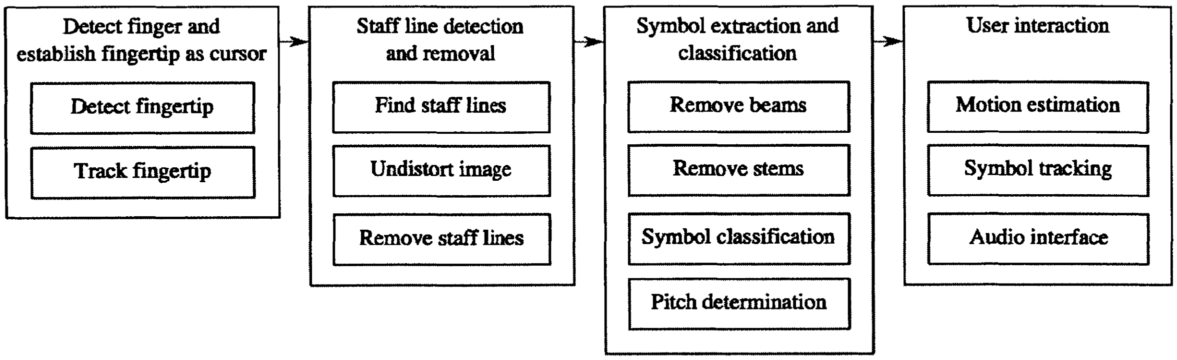


Figure 6-2: The processing and interaction pipeline in our assistive music reading system.

page, that in most cases includes but one or two bars, rather than the entire page as in traditional OMR systems. While solutions to interpret partial views of the page were suggested, our approach is using a finger-steerable camera that is meant to be controlled in a sightless way. On one side this creates an opportunity, as a local view can be easier to computationally analyze, however on the other side this introduces additional problems to tackle such as: using the finger as the cursor for the analysis, handling a moving view of the page and providing feedback on the scanning operation itself rather than just the content (the musical symbols). These issues augment the existing necessities of a traditional OMR pipeline that also appear in our system: staff line detection and removal, segmentation, classification and more.

Our computer vision system has three operation states: Detecting finger, Detecting staff lines, and Reading symbols. The states are organized in a simple state machine that advances with the goal of reaching the "Tracking Notes" state, which implies that a finger, staff lines and notes were detected and are currently being tracked in the scene (see Figure 6-3). Each state has a different audio feedback which informs the user of the status.

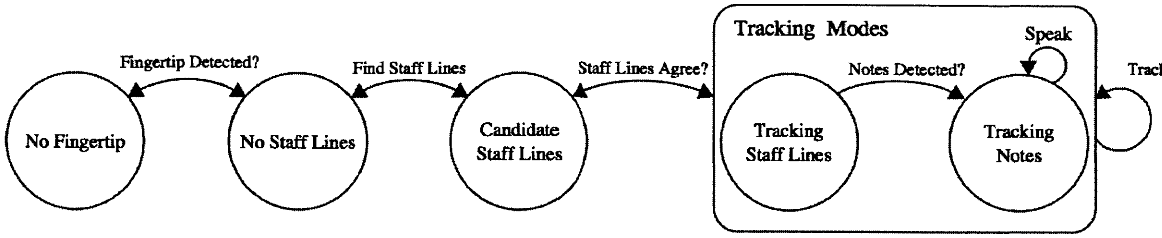


Figure 6-3: The processing pipeline state machine.

6.3.1.1 Fingertip Detection and Tracking

Finding the fingertip in the image is imperative to the operation of our system, as we look to create a strong non-visual mental bond between the finger and the underlying printed music sheet. We therefore look for the fingertip to achieve the following goals: boost performance by constraining our symbol recognition efforts, provide feedback on symbols directly in front of the finger, and guide the users to bring their finger to an optimal position for reading a line of symbols.

Fingertip detection is done similarly to the Mobile-FingerReader (see Section 5.3.3). The input image from the camera is converted to the normalized RGB (nRGB) colorspace. The nRGB is useful for coping with varying lighting conditions, and also for detecting skin color in the nR channel. The nR channel is binarized using an adaptive threshold, and examined for a connected component that emerges from the bottom, seeing as the camera is looking at the paper down the direction of the finger. The upper-most point on the boundary of this connected component consistently correlates with the fingertip. We use a coarse-to-fine approach to boost performance, by first considering a downscaled binary image to extract a rough fingertip point, and then in the original resolution image we only examine a small region of interest around the rough estimate. The resulting raw fingertip estimate is fed through a standard Kalman filter with a 4-entry state vector (2D position and velocity) to smooth measurement noise with error covariances that were determined empirically. See Figure 6-4 for an illustration of the fingertip detection and tracking process.

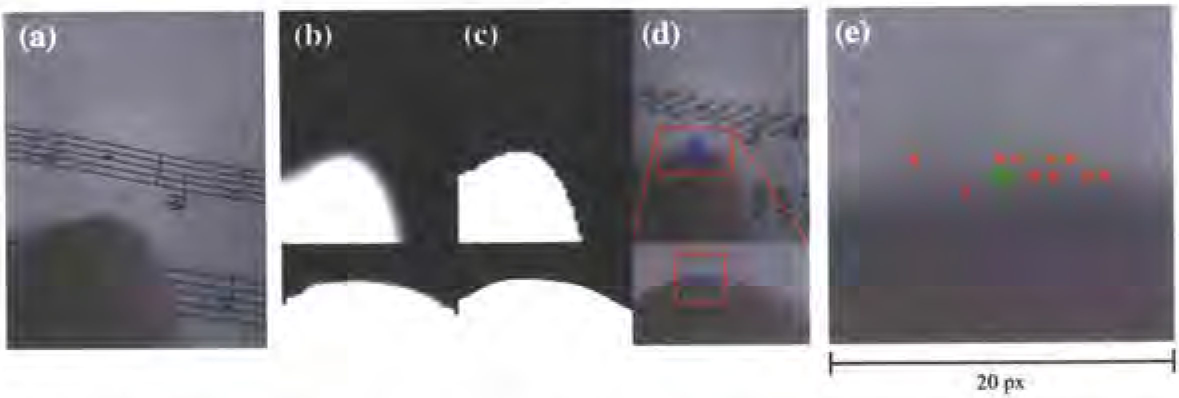


Figure 6-4: Fingertip detection process. (a) Original image, Coarse-to-fine detection of the fingertip: (b) detected skin probably map, (c) binarized using Otsu method, (d) detected fingertip point, (e) Kalman filtering of the measurement point (green - filtered result, red - measurements with noise).

6.3.1.2 Staff Lines Detection, Removal and Tracking

The fingertip location in the image allows us to consider only a small region of interest where we look for the musical constructs, which leads to a short running time. Within this region we look for the staff lines.

We use an adaptation of the Roach and Tatem method for staff line detection [186], in which they look for longest horizontal black pixels run in the image. Assuming the staff lines run from the left extremity of the image to right, we pick black pixels on the left and match them with black pixels on the right, and maintain the lines that have the most black pixels along them. We look for a best candidate for each staff line by using a sorted set based on the line's intercept. The set operation prevents us from choosing more than a single line for a given intercept (with a small threshold), therefore in the end of this operation we are left with 5 unique staff lines. For validation we calculate the distance between neighboring pairs of lines and their angles, and deem the line detection a good one if all measurements agree within a small variation. Disagreeing detections will prevent the system from going on with the rest of the OMR pipeline. For further calculations we extract the staff line space

(SLS) and staff line height (SLH) from the detected staff lines.

In subsequent frames after staff lines were successfully found, we search for new lines only within a small region around the previously found lines, to speed up computation. The lines are tracked with a Kalman filter, so intermittent lapsus in earlier stages of detection that hide will not break the continuous operation, rather the system will strive to recover within a given number of frames.

6.3.1.3 Undistortion

Imaging from the finger results in a close-up view of the page, and also creates a distortion of the paper plane in the image due to the angle of the finger in relation to the page, and the perspective camera properties. This distortion may be negated (discounting for the radial distortion caused by the camera lens) by applying the inverse projective transformation between the page plane and the image plane, which in turn can be calculated if we knew the relative position of the camera or otherwise had prior knowledge of the print. Since it is a requirement of our system is not to have prior knowledge of the print, and in most cases there are insufficient features in a single close-up image to robustly recover the angle (e.g. lines with a vanishing point), we cope with the distortion using 2D rigid transformations. This simple and fast undistortion technique proved to be effective, and improved our symbol recognition rates greatly.

The detected staff lines impose a near-uniform 2D rotation of the image, although in some cases, depending on the perspective distortion, the lines disagree on the angle. Using the inverse 2D rotation roughly rectifies the symbols for proper classification. All classification operations are then performed on the rectified image.

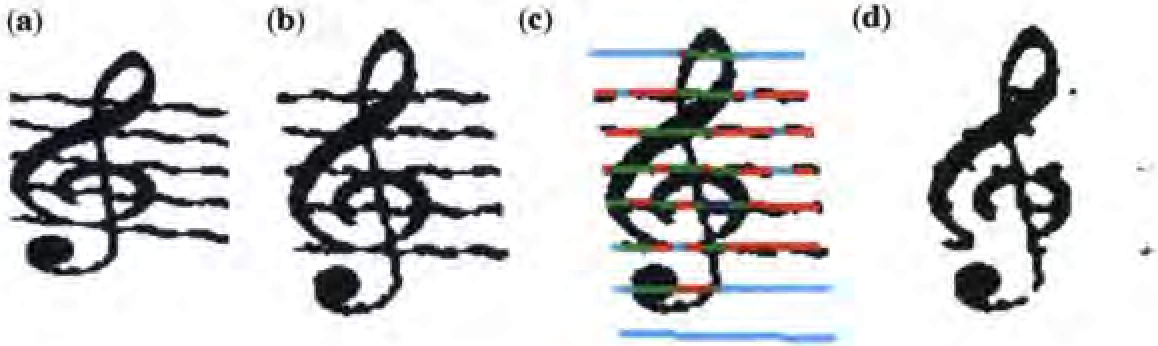


Figure 6-5: Classification of staff lines for removal. (a) the original binarized image, (b) the undistorted view, (c) annotated staff lines from the HMM, and (d) the staff lines removed.

6.3.1.4 Staff Line Removal

For removing the staff lines and keeping the musical symbols intact we use a Hidden Markov Model (HMM). We sequentially scan the staff line to determine at each point whether it belongs to the staff line or a symbol. The hidden states we utilize are: { STAFF, SYMBOL, SYMBOL THIN, NOTHING}. The STAFF state models a staffline segment, the SYMBOL and SYMBOL THIN states model a symbol segment or a thin part of a symbol, and NOTHING models segments that not staff or symbol. The sequence of observations is created from a scanning of a staff line pixel-by-pixel, and counting the number of black pixels above and below the line. The HMM alphabet is comprised from 16 symbols using a 4-bit code, where every 2 bits represent the count of black pixels in each side. Transition and emission matrices were manually calculated from a number of annotated examples. To discover the annotation for a new staff line we calculate the observation sequence from traversing the line and run the Viterbi algorithm. The resulting hidden states sequence gives us the annotation for the staff line, which then we use to remove all pixels in the STAFF or NOTHING state according to the staff line height (SLH). We used [193] for the HMM implementation.

6.3.1.5 Symbol Detection and Classification

Having removed the staff lines we are left with the exposed musical structures, however we apply a number of additional operations before obtaining a full classification of type, pitch and duration. We begin with cleaning operations: removing beams that connect notes and removing stems. After we are left with note heads and other symbols (such as accidentals, slurs or dots) we classify them based on their geometrical properties using a dataset of examples. Once classified we determine the pitch for accidentals and notes, and optionally look for duration marks such as flags, beams and dots. Finally, we audibly report to the user on the extracted notes in the image.

6.3.1.5.1 Beam Removal To correctly classify beamed groups of notes, we must remove the connecting beam. In many cases the beam reaches outside of our small region of interest, making it harder to distinguish between a beamed note and symbol that is not yet fully in view. Therefore, beams, if any such exist, are removed first so to make the rest of the pipeline agnostic to their existence.

Our beam-finding algorithm is based on an interval run-length encoding. First we detect vertical segments in the binarized image that are likely to be part of a beam. We look at a run-length encoding of a single column and keep the segments that are in the [SLH + 1.5, SLH * 4] range (SLH being the Staff Line Height). Thereafter we look for a consecutive overlapping segments whom centers also converge on a line, since a beam is always a straight thick line. We finally remove the selected beam segments by painting over the pixels. This process then works nicely with our staff detection algorithm, and is illustrated in Figure 6-6. The position and existence of the beam is recorded in order to determine if a detected note is an eighth note or a quarter note. This method is also effective in separating connected eighth notes into individual components.

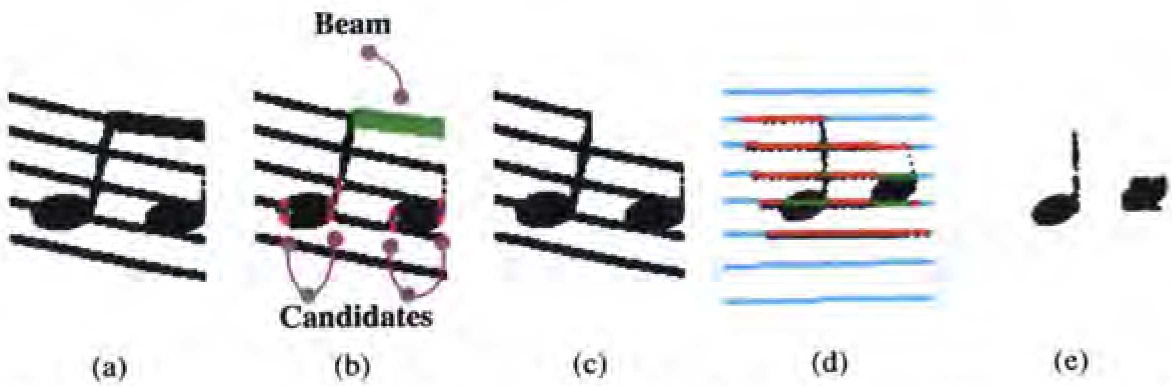


Figure 6-6: (a) the binarized input region, (b) detecting consecutive segments of viable height, (c) removing the segments classified as beams, (d) staff line removal on the rectified image, (e) result with beam and staff lines removed, ready for note classification.

6.3.1.5.2 Stem Removal Many of the note symbols arrive at this point of the OMR pipeline with a stem (the vertical line going above or below the note head): half notes, quarter notes and also eighth notes after having their beam removed. In order to make a simple classification step we discard the stem and then consider only the note head for classification. To remove the stems we look at the rectified image after beam and staff line removal. We perform a reduce operator on the y-axis to get a projection and study the derivative (via a 1D Sobel operator) of the projection vector. Local extrema of the derivative correlate with points where the thin stem transitions to the thick note head (see Figure 6-7). We filter out unlikely breakpoints, e.g. points that are too close together or too close to either end of the symbol.

6.3.1.5.3 Symbol and Pitch Classification Once we obtain a clean symbol we use geometric features of the contour with a decision tree classifier to classify the symbol to its type (e.g. note head, accidental, bar line, etc.). Inspired by [177], we use the following features: width, height, area, ratio of black vs. white pixels, and 7 Hu moments. All features are taken after a normalization according to the staff line spacing (SLS) calculated earlier. The decision tree is set to have a maximum depth

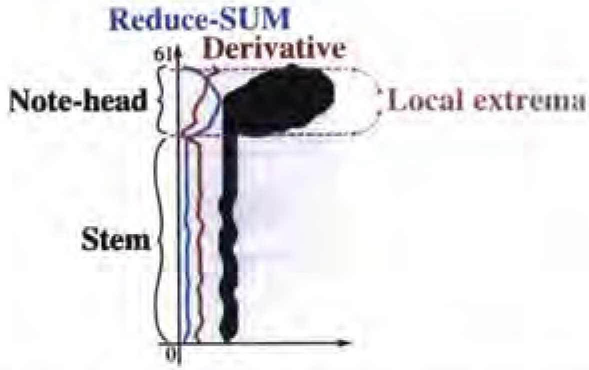


Figure 6-7: Stem detection via projection and derivative extremum.

of 20 and at least 2 samples in a node, and trained over a dataset of 1170 manually classified note symbols from a training set.

To determine pitch for relevant symbols we obtain the central point of the symbol, for note heads we consider the the center of mass and for incidentals (sharps and flats) we geometrically determine the central point from the height and width of the contour. Distances of the central point from the staff lines (and ledger lines) are calculated and the closest two distances are retained. If the closer distance one of the two is less than 15% of the SLS, we deem the symbol to be on the staff line, and if the two distances are roughly the same (and less than 75% of the SLS, to eradicate errors) we deem the symbol to lie between the two staff lines. We thereafter assign an octave and pitch to the symbol based on the two staff lines. See Figure 6-8 for an illustration of this process.

6.3.2 User Interface and Feedback

In order to facilitate a scanning operation for a VI person, we rely on the previously developed mechanism for scanning text in the FingerReader (see Chapter 4), with modifications to support reading stave notation music. The feedback is divided to two main parts: scanning feedback via audio tones, and music notation feedback via

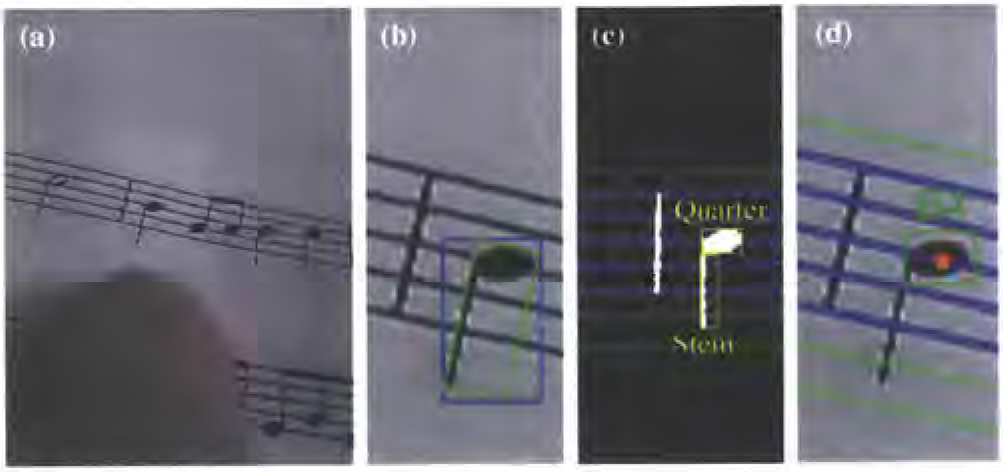


Figure 6-8: Symbol pitch and type classification process: (a) original image with region of interest, (b) detected contour to classify, (c) the segmented contour of note-head and stem, (d) the classified note-head based on the position on the staff lines.

speech.

6.3.2.1 Speech Feedback

Following the works on translating musical notation to speech [37], each note is translated to duration and pitch. For a note the system first utters the duration: "eighth", "quarter", or "half", followed by the pitch class in latin letters (CDEFGAB), and finally the octave ("3", "4" or "5"), for example "eighth-D4". For accidentals the system utters the class and pitch (e.g. "Flat-D4"), and for symbols without pitch the system simply utters the word (e.g. "bar", "quarter rest").

Since the notes are digitized, they could just as well be played out in their appropriate pitch with a virtual instrument (a MIDI piano for example). We decided however not to implement this feature, since the goal of this system is to recreate the note reading experience, which does not include playing the notes aloud but rather mentally reconstructing them.

6.3.2.2 Tonal Feedback

The tonal feedback guides the user in scanning a line of stave notation music. The goal is to help the user keep the finger-camera pointing at roughly the middle of line, via a feedback that describes the distance from the center of the line. The major difference from the FingerReader is that our experiments with a blind user revealed it is much easier to follow only two feedbacks at a given moment: above the line (a high C note), and below the line (a low C note), instead of a continuous varying tone to describe the distance itself (as described in 5.4.4). This simplification greatly reduces cognitive load in operating the device, as the user gets no tonal feedback if they are roughly centered and can concentrate on the speech feedback. When the system cannot detect any line in the image it emits a G note tone.

6.4 Evaluation With Blind Musicians

To evaluate the performance and usefulness of the system we performed a controlled user study with VI musicians. The goal of the study was to assess the feasibility of the MusicReader to assist in reading a printed music sheet in an unstructured environment, simulating the real situation a person would wish to use the device.

6.4.1 Participants

We recruited 5 participants (4 male, 1 female, aging 18 to 33 years) from a pool of volunteer VI musicians. Table 6.2 shows the details of our participants in terms of visual impairment and printed music access habits. An additional VI musician volunteered to act as a pilot for the user study to reduce bugs in the system and

Participant	Age	Gender	Hand	Visual Impairment	Music Access
Pilot	33	Male	Right	Totally bind. Onset at age 3.	Music braille. Learn by ear.
P1	24	Male	Right	Retinal detachment. Able to see some color and light.	Music braille. Learn by ear.
P2	18	Male	Right	Retinopathy of prematurity.	Human transcriber.
P3	28	Male	Right	Anophthalmia. No residual vision.	Learn by ear. Music braille.
P4	19	Female	Right	Leber hereditary optic neuropathy. No residual vision.	Music braille. Learn by ear.
P5	27	Male	Right	Totally blind.	Human transcriber.

Table 6.2: Participants in our user study.

the testing procedure (in Table 6.2 appears as "Pilot" participant). The participants were compensated for their time and effort.

6.4.2 Procedure

The study consisted of four central parts: Pre-usage interview, Training session with the reading device, Reading tasks, and finally a Post-usage interview. The whole session lasted 60 minutes on average per participant. The pre-usage questionnaire contained questions about current habits of reading printed music sheets (reported in Table 6.2) and the major problems with these techniques, as well as the needs and wishes from a technology to assist in that domain (reported in the Introduction and Discussion sections).

The training session (10-15 minutes) introduced the concept of reading music with the finger, listening to the audio cues to track a line of a musical sheet. The training sheet contained simple C-major scale hikes and arpeggios in different note duration (half, quarter and eighth). During training we provided the readers with constant help in listening and using the audio cues for tracking as well as positioning the hand correctly, however we did not offer such help during the actual reading tasks.

After training the users were given two printed music sheets to read, with 56 and

Figure 6-9: One of the sheets the participants used for reading, with arrangements of "Happy Birthday" and "Greensleeves".

62 notes on them (16 and 14 eighth notes, 18 and 26 quarter notes, 22 and 18 half noted, and numerous accidentals and rests) in 3 lines of standard staff notation of well-known melodies. The melodies were simple arrangements (no harmony) of the following standards: "Happy Birthday", "Greensleeves", "Over the Rainbow" and "Amazing Grace" (see Figure 6-9). Participants were given up to 20 minutes to try and read the two sheets, and questioned about whether they can recognize the melodies in them. The read notes and audio feedback events were recorded by the software on the PC along with timestamp, and were later analyzed.

To gather qualitative feedback we performed an exit interview with a number of Likert scale questions as well as a semi-guided open discussion about the general usage experience. We took notes during the interview and video-recorded the sessions, which were later transcribed, coded and categorized. The results of these interviews are reported in the Results (6.4.4) and Discussion (6.5) sections.

The users were seated at a table with the sheets of music in front printed on regular paper. The MusicReader device was placed on their finger attached to a laptop computer placed on the table as well (see Figure 6-10).

Figure 6-10: User study participants in midst reading a music sheet.

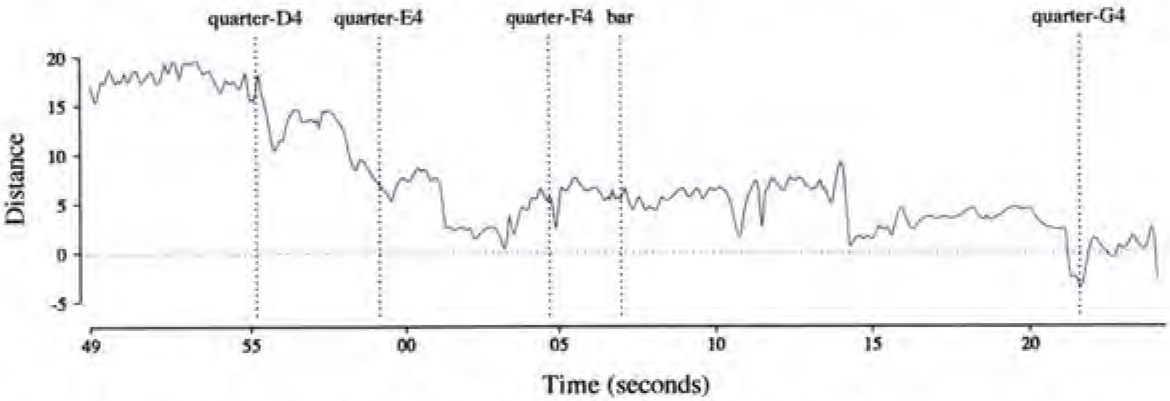


Figure 6-11: Time-series data recorded by the MusicReader software, a 35 second excerpt from P1's reading session. Distance is measured in pixels from the center of the staff lines.

6.4.3 Analysis

The music reading software system records a stream of two types of events: Note found, and Distance from line, where each event has a timestamp (see Figure 6-11 for an illustration). To analyze the proficiency of a participant in reading the printed music we take a similar approach to Section 5.4.6 where we look for longest intersecting subsections between the extracted notes and the ground truth (see Figure 6-12). Then we construct a histogram that counts the number of occurrences a certain length subsection was read, and assigns a score for each bin of the histogram.

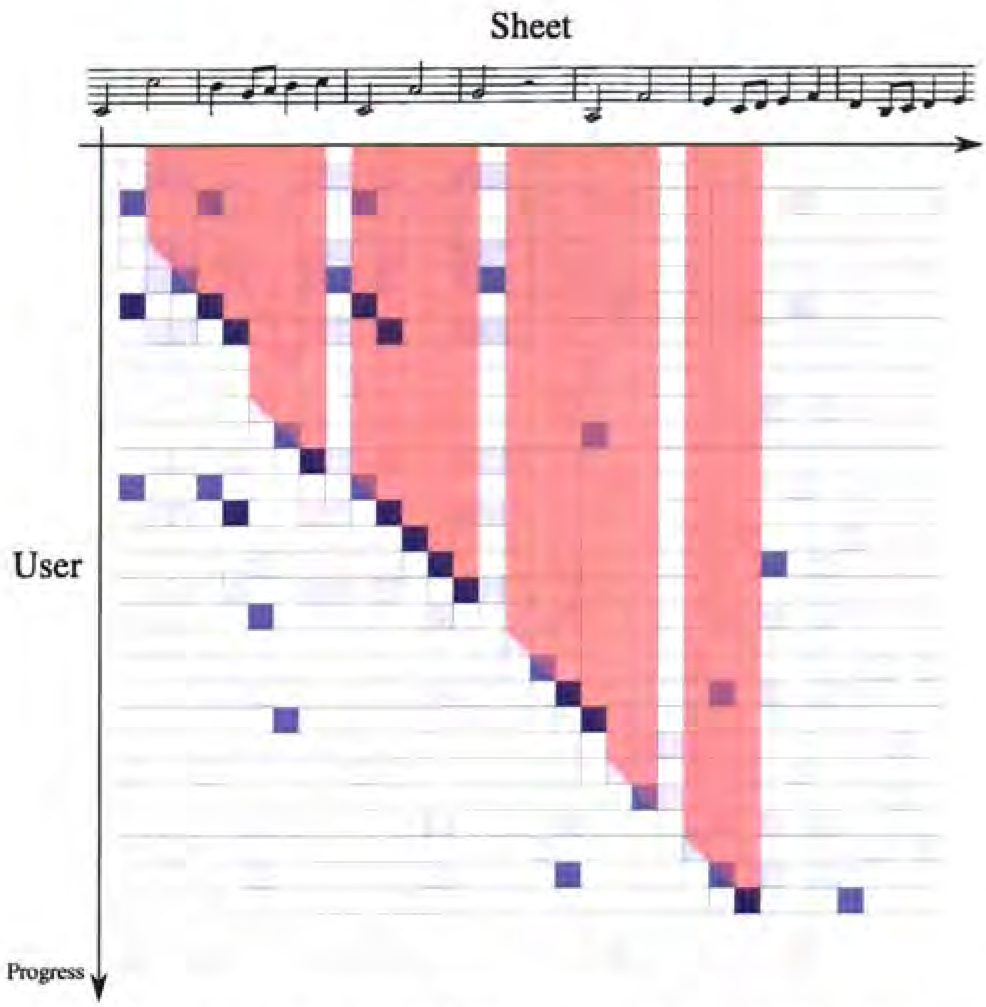


Figure 6-12: Match matrix between the ground truth (music sheet, columns) and the user's scanning. Blue color represents a match between a note in the sheet and a note the user read. More saturated blue represents a longer sequence of matching notes. The red area helps visualize what relative portion of the sheet was read. The matching is performed using a dynamic programming algorithm: a modification of the standard text edit-distance.

The blue squares not on the leading diagonal (marked red) represent misalignments in the matching algorithm. Either the algorithm found a shorter, partial match at a different time of the scanning process (e.g. re-reading, skimming), or a different section of the sheet. The marked diagonal represents a continuous reading state, where the user was able to consecutively extract many notes from the sheet in sequence.

The "Total Words Read" method of the Mobile-FingerReader does not apply easily in the case of reading music. There is a far smaller subset of "words" our system can read: 2 octaves of possible notes in 3 durations (Eighth, Quarter and Half) that repeat very often throughout the test pieces. Measuring the reading success in terms of percentage of read notes will therefore not contain as much information as a measurement that models reading notes in a sequence.

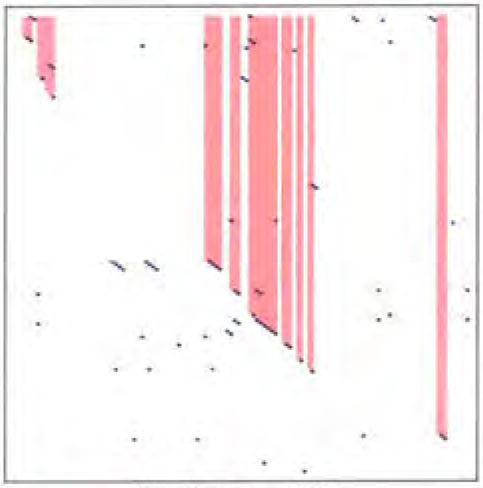
6.4.4 Results

We report on our finding from the study in three parts: quantitative measurements, qualitative measurements and impressions from the study alongside issues raised in the post-usage interviews.

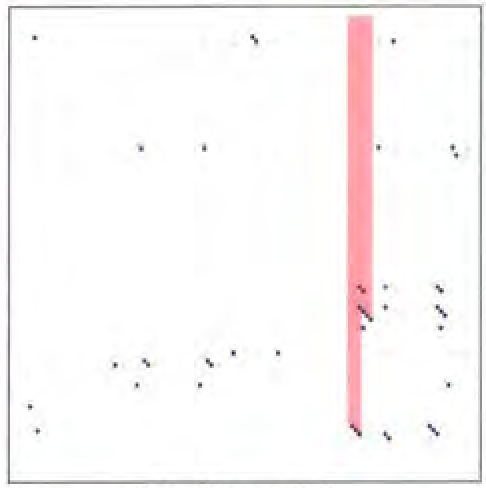
6.4.4.1 Quantitative

In Table 6.3 we show the results from the analysis of reading proficiency in terms of consecutive extracted notes. Participants P3 and P4 were more successful at reading compared to the others, while P2 received the lowest score, which is commensurate with their qualitative reporting of efficiency and enjoyment (in Table 6.4). While the results show the users were able to extract only %10-%50 of the notes from the two test sheets, all users were stopped by the examiner at the time set for reading independently (20 minutes). Given additional time, users would continue to read more notes.

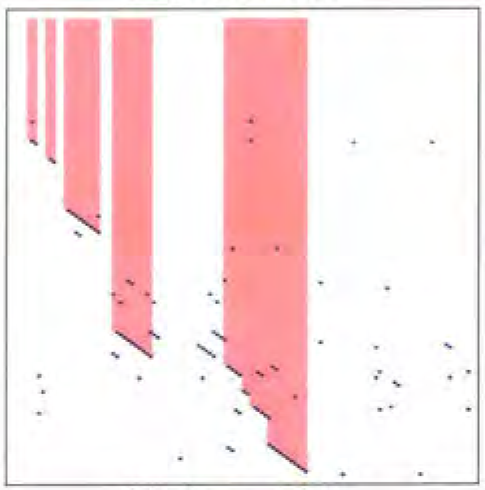
The score scale is therefore arbitrary and meant only as a comparative tool between users. It does correlate however with the qualitative response on enjoyment, which suggests it could be used to model one's success in using the system. Due to the small sample size we did not perform deeper statistical analysis with the goal of generalizing on these results.



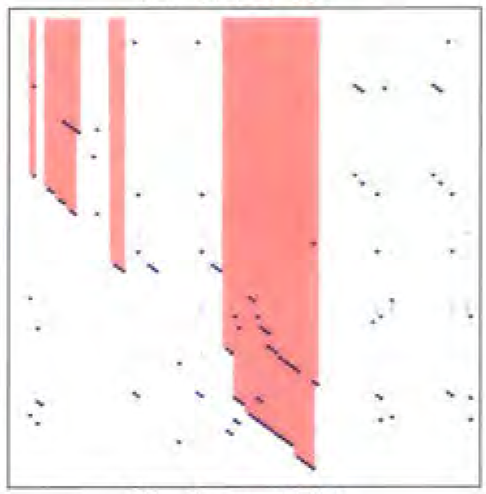
P1. Score: 1285



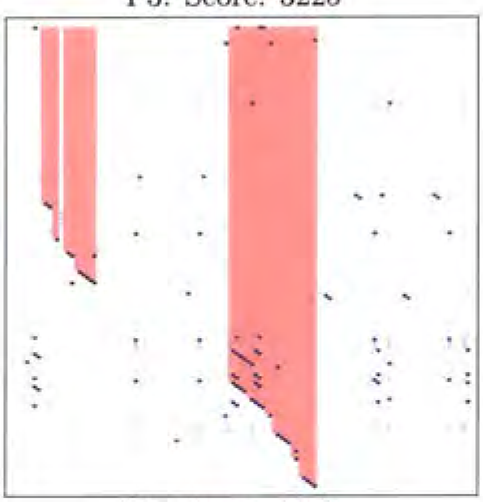
P2. Score: 557



P3. Score: 3225



P4. Score: 2729



P5. Score: 1846

Table 6.3: Scoring and matching matrices of the participants.

6.4.4.2 Qualitative

In Table 6.4 we list the ranking from the participants on the Likert scale questions. Overall, participants found the experience to be somewhat enjoyable, however not particularly easy. Participants were perceived to do quite well in following and understanding the feedback, and to a lesser degree in recovering most of the the notes and recognizing the music.

Participants did not think the MusicReader was easier than other music reading aids, although in the interviews some participants reported that they do not know of similar aids. Most participants (all save for P2) stated they would require an expert to help them operate the device but felt there weren't many things to learn. To a degree, users agreed they now feel greater desire to read printed music independently, however not necessarily with the MusicReader.

In the post-usage interviews participants spoke freely about their experience using the MusicReader, while the investigators encouraged them to reflect on different aspects of their experience. The following is a synthesis of the major issues raised in the interviews, combined across the participants:

- **Would like to know the bar number or location in the piece**: A number of participants (P1, P2 and P4) requested the ability to know where they are in a big piece, a thing they marked as important when playing with other musicians. They wanted the computer to keep track and report of what bar they are on, by speaking out the bar ordinal number.
- **Reading process was slow**: Other participants (P1, P3, P4) reported they would require a faster rate of reading if this is to be used in a real life scenario. Both P4 and P3 mentioned a rehearsal situation where the musician must be as fast as the conductor to keep up with the band. P1 however related the time-

		P1	P2	P3	P4	P5	Avg.±Stdev.
	General						
Q1.	The overall experience was enjoyable	3	4	2	3	2	2.8 ± 0.75
Q2.	Accessing Music with MusicReader was easy	3	5	4	3	4	3.8 ± 0.75
	Difficulty / Success						
Q3.	I was able to recover most of the symbols	3	3	3	2	2	2.6 ± 0.49
Q4.	I was able to understand and recognize the music	3	3	4	1	3	2.8 ± 0.98
Q5.	I was able to follow the audio feedback	2	1	1	1	2	1.4 ± 0.49
	Comparison to other music reading aids						
Q6.	Accessing music felt easier with the MusicReader	4	4	5	4	3	4 ± 0.63
	Independence						
Q7.	I would need the help of a technical person to be able to use the MusicReader	1	1	5	2	3	2.4 ± 1.50
Q8.	I felt I needed to learn many things before being able to use the MusicReader	2	2	5	4	4	3.4 ± 1.20
Q9.	I feel greater desire to be able to read music independently	3	2.5	1	3	2	2.3 ± 0.75
Q10.	I desire to use the MusicReader to access music on the go	3	3	3	3	3	3 ± 0.00

Table 6.4: Questionnaire results from users of the MusicReader. 1 - Agree, 5 - Disagree.

consumption problem to the former item, where orienting himself in a piece will simply take a very long time.

- **Feedback modalities - Tactile**: Two participants (P3 and P4) wished there was another output modality in form of a tactile feedback. P3 asked for vibration feedback to help with aiming the finger at the text, however P4 thought the vibration is better to provide the line-tracking signals. P4 suggested to add a refreshable braille display to output the information from the page instead of speech.

- **Feedback modality - Audio and Speech**: P2 mentioned that providing feedback about music in form of musical notes is confusing, if one is trying to mentally reconstruct the music while also listening to notes in various pitch, and perhaps a different octave should be used. P2 and P5 both requested a more appealing synthesized voice output as they thought the current one is too robotic, and P3 mentioned playing back the note instead of speaking it out could assist.

- **Accuracy problem**: Two participants (P1 and P2) identified a problem with the system's accuracy. They stated that in a real life situation there is no chance to go back and re-read a section if they believe it was improperly read, as they are required to read fast and play right away.

- **Reading more from the printed sheet**: Participants P1 and P2 noted that a sighted person can easily find the key and time signature as soon as they see the printed sheet, and this is imperative for correctly playing the music. P3 requested the ability to read the lyrics that often accompany the stave notation music on the page.

- **Success and enjoyment**: All participants agreed the MusicReader is an interesting concept that, if further developed, could allow them to access music in situations where today they cannot do so (specifically band practice). P5 and P2 noted the design of the device is comfortable, easy to move and mobile.

- **More types of feedback**: P2 and P4 would have liked the device to help them with aiming the finger, and P3 requested the ability to review what was already read with the device. P5 suggested a dual-step operation could assist by first giving a higher-level overview of the paper and only then zoom-in to start the reading process.

6.5 Discussion

The MusicReader is a novel approach in the domain of mobile assistive technology for VI musicians, where most prior work did not attempt to tackle non-visual reading of printed music. As such, our study participants had mixed comments about its utility, although there was a positive consensus about the potential it has. The following is a discussion of the key findings rising from the study as well as the work on the MusicReader.

6.5.1 Learning-by-ear

While most of the participants in our experiment noted the MusicReader is an intriguing technology that could be useful if further developed, all participants agreed that learning-by-ear is still the best tool they have to access music.

> P3: "I would love to be able to read music, but I still consider having aural skills, the ability to learn a piece by ear and play it back, [to be] a very important tool. It could be used in combination with reading, and I still think being able to read is a good thing."

P1: "Even with a technology that could tell me what is on the paper I don't know if I would prefer it to just listen and figure it out."

On the other hand, our interviewees reported of numerous situations where learning-by-ear is impossible or impractical: band practice, working with a conductor, in the classroom and while teaching. In these situations, according to our participants, VI musician are at a disadvantage even in spite of their technical abilities.

P3: "Not being able to access printed music material knocks blind people out of a big segment of the market. I don't think I could go audition for the BSO [Boston Symphony Orchestra], even though I think I have the chops to at least play 3rd or 4th trumpet for the BSO, but they want you to be able to sight-read. Not to be able to work on-the-fly like that is a really big problem. [...] At the moment I would not be able to teach beginners that don't know how to read, but I can certainly teach them how to play. I feel like that's something that keeps me from teaching beginners privately."

P1: "[in a situation where] someone hands out chord charts for a song or a melody that I would need to learn, I would still need to memorize it, but [being able to read it] would be helpful [...] in any sort of band that you play for or classroom. If I had to teach a blind person I wouldn't use a printed chart."

6.5.2 Finger positioning and aiming

Most study participants had problems of aiming the device and maintaining the right angle for proper reading. This was reported also in the FingerReader and the Mobile-FingerReader. This problem in the MusicReader is even more acute as VI musicians read with the specific goal of playing their instrument, and therefore can spare, at most, one hand for reading depending on the instrument they play.

> P3: "I can't have my left hand off of the trumpet, I must hold it. [...] I have to have both hands on the instrument"

Study participants were also concerned with getting a very quick and precise reading, and did not have much patience towards learning the hand positioning or maintaining it for long. We conclude that both the imaging hardware as well as the software may need to improve to overcome this problem. The camera lens could be of a wider angle, and the algorithms to find the fingertip and staff lines could have a much higher tolerance towards skewed views.

> P4: "Aiming was hard. Maybe there should be some sound, vibration or speech to tell you if you're aiming right. If you're angled too far left or right you'd get a vibration, and it would click if the finger is positioned right."
>
> P2: "[...] the finger needs to be in a very specific position, there should be a better way. The angle was not directly straight with the paper, and I can't see the paper."

6.5.3 Tonal signals for reading music

Some participants reported of an increased cognitive load when listening to the assisting tones while trying to mentally reconstruct the music only from the spoken names of the notes. Additional research is needed in order to measure the added cognitive load, as simple measures already implemented in the system (see Section 6.3.2) are insufficient. In reading music with the MusicReader this issue of mental interference is more severe than in reading words with the FingerReader, which suggests tonal feedback may be less effective. Participants suggested we incorporate tactile feedback to circumvent this issue.

> P2: "The notes are good for feedback, but if you're thinking about the music - that's confusing. Maybe it shouldn't be in the music range, not C,G and C if I am reading something in a C scale."

6.6 Conclusion

This chapter presented the MusicReader, a finger-wearable camera that assists visually impaired musicians in reading printed music sheets. The system includes an Optical Music Recognition (OMR) pipeline that recognizes multiple types of stave notation musical symbols in real-time speeds, focusing on a small region around the fingertip. We presented the technical details of the system as well as a user study with 5 VI musicians that established the need for mobile assistive technology to assist in reading printed music. Study participants revealed a number of situations where not being able to read printed music leaves them at a disadvantage or impeding their advancement, even if they are confident in their playing skills.

The findings from our study point both to the potential of the MusicReader as a mobile assistive technology and to the usability obstacles of such an approach. Reading printed music for VI musicians is not a special case of reading printed text but rather a new problem class. In many situations music is read with the goal of immediately playing it, often in a group setting with other musicians, which requires a fast response, high accuracy and less than ideal reading conditions. Reading music also requires the reader to mentally reconstruct the music, which can interfere with any audio feedback from the system. Nevertheless, some elements of reading music are similar to reading text such as locating oneself in the page.

We look forward to improving the MusicReader to enable a better user experience in terms of accuracy, speed, tolerance and feedback. These issues are prime candidates for future research into the domain of assistive mobile printed music reading, involving both engineering efforts and further user studies. We also believe the MusicReader could be used by non-VI musicians, for example people with dyslexia or music learners.

This chapter concluded the part of this document that discusses the contributed concrete work to finger augmentation. The next chapter will provide the reader with an overview of other major contributions, ideas and concepts presented throughout the thesis document, as well as an outlook to future research directions finger augmentation can take.

Chapter 7

Conclusion

This dissertation covered my work on *Finger Augmentation*, creating finger-worn active devices for assistance. This concluding chapter will reiterate the contributions and lessons learned rising from the four prototypic explorations I've created and presented in Chapters 3, 4, 5 and 6, as well as the survey and classification framework of previous works presented in Chapter 2. The next section presents a chronological account of the work over the last five years in developing finger-worn cameras. The following sections discuss the benefits *Finger Augmentation* provides to scaffold manual interaction, as well as possible future directions of research including the particular type of augmentation I focused on - using finger-wearable cameras.

7.1 Augmenting Fingers For Assistance: Reflecting on the Last Five Years

Our earliest work, the EyeRing (Chapter 3), was not the very first finger-worn camera, however this statement depends on the definition one takes for a camera. The Symbol company recognized the potential of using finger-worn laser imaging to scan barcodes by pointing at least as far back as 2000 (when the product was released) [232], and Ando et al. soon followed with a similar exploration of assistive applications [6]. A few more investigations of this idea were made between the early 2000s and the EyeRing (2010) [142; 263], however high-resolution imaging and computer vision were not part of their agenda. Since its inception the EyeRing sparked in the community a keen interest in this domain [152]. Ideas ranged from being similar to the EyeRing [184] to utilizing completely different camera positioning, hardware and applications [257]. The EyeRing left the space mostly unexplored, and only sent one probe outside the core assistive motif to examine finger augmentation for office work [153]. That void was soon filled by my own and other people's work.

The EyeRing's focus on assistive applications for people with a visual impairment determined the direction of the next major project - the FingerReader (Chapter 4). At the time we started contemplating the idea that would become the FingerReader (circa the end of 2012), before there was any hardware put together and the software was minimal, it was obvious we uncovered an interesting and simple idea. Using the pointing gesture and the index finger to guide a high-resolution camera, especially in the context of visually impaired users, was well received in our colleagues in VIBUG (a group of visually impaired technology users in MIT) and in the HCI community [204]. This concept, while far from new, was reborn by the idea of using the finger to read - to level the playing field for seeing and visually impaired people.

The success of the FingerReader, which was manifested both in the media and academia [205], demonstrated the power of the idea of reading with the finger. This reconnection of the traditional usage of the finger to trace lines of text with the assistive tradition of reading braille with the fingertip, was immediately understandable to anyone. In the aftermath of the media coverage (that peaked in February and July of 2014) we were bombarded with requests from people and organizations who wanted to receive a FingerReader device to use or test. We collected over 3000 email addresses from people around the world who were interested in hearing more of the project as it progresses, and the emails are still trickling in to the time of writing these lines. What was equally surprising and incredible, are the approaches from people to get a FingerReader not for a visually impaired person, rather for someone with dyslexia, brain-related disorder (e.g. patients recovering from stroke) and people interested in immediate translation. This was the seed for the *Assistive Augmentation* spectrum and design principle: assistive technology crossing over to other target audiences.

In the CHI 2014 conference we held a workshop around Assistive Augmentation to explore this concept with researchers both in assistive technology and human augmentation [76]. We were looking to find a reconciliation between the two concepts, one very established and one still young and finding its way. Our participants showed demos ranging from a system for sensory substitution [129] to an ear-augmenting input device [124], and engaged in lo-fi prototyping of devices that could be used both as assistive technology as well as by other communities. The fluidity of a technology's purpose was not a new idea, as Yvonne Rogers eloquently writes: "technologies are *interpretively flexible*" ([187, p. 71]), however this was not thoroughly tested in the context of assistive technology. The idea of the FingerReader was powerful enough to drive people to imagine its benefit beyond what we ever thought of or intended for it, which was a perfect example of an Assistive Augmentation technology. The notion of a connection between a visual impairment and dyslexia in the domain of reading, of which Challis wrote in his work on assistive devices for reading music [28], drove us to further understand the breadth of the concept.

Later in 2014 we began to work on extensions of the FingerReader to further develop the prototype and also venture out to reach other communities. One direction was making the FingerReader more widely accessible by getting it to run on a standard mobile phone device - the Mobile-FingerReader (Chapter 5). People we worked with in testing the FingerReader stressed that the best usage for this technology is not a stationary application but rather a mobile one. Results of a user study with the Mobile-FingerReader shown good potential for a viable product, so it is now undergoing a process of productization with manufacturers in China.

The second descendant of the FingerReader was the MusicReader (Chapter 6). Growing out of the Assistive Augmentation concept, we identified a prospect target audience for finger-based assistance - blind or visually impaired musicians. We began a collaboration with Berklee College of Music towards the end of 2014, in parallel to our efforts on the Mobile-FingerReader. The collaboration quickly bore fruit as we prototyped the MusicReader system based on the FingerReader PC-based platform and tested it with students of the Assistive Music Therapy in Berklee. We discovered almost an open field of opportunity for developing this kind of technology, as Optical Music Recognition - OMR (in contrast to Optical Character Recognition - OCR) was virtually not applied to build assistive technologies. Most lessons learned from the FingerReader applied for the MusicReader in terms of non-visually guiding a person to read a printed line, however the music application had its own needs. While users of the MusicReader saw a potential solution for certain situations, they agreed that learning by ear is still the best way for them to learn music.

7.2 Augmenting Hands For Assistance: A Parallel Track

While working on finger augmentation, I also had the chance to make explorations in augmenting hands through smart tools. An endeavor that started from a collaboration with Amit Zoran on his FreeD device [270; 269; 268] escalated to a full-blown research project: the Digital Airbrush. While out of scope for this thesis, the Digital Airbrush bears resemblance to the works presented here in the sense that provides real-time manual assistance to the user. The goal of the Digital Airbrush is to support an artistic process for a novice spray-painter and not overcome an impaired sense, although one may think of an undeveloped sense of drawing as an impairment.

We used a GREX Genesis.XT pistol-style airbrush, relieved of its rear paint-volume control knob and fashioned with a custom made augmentation of our design. To the airbrush body we added: a 6DOF magnetic tracker, a mechanical constraint on the trigger, a potentiometer (POT) on the trigger, a 6DOF inertial measurement unit (IMU) and an LED. The trigger constraint lets the painter to only spray as much as the software allows at a certain point in space.

We implemented a GPU-based control algorithm to determine if the painter is in risk of spraying in the wrong direction and location, as calculated from the information from the tool, and issue control commands to the tool at ~100Hz. The system simulates in real-time the paint deposit based on a parametric model of the typical spraying pattern of the airbrush, with respect to distance from the canvas, the trigger action and the radius from the center of spray projection.

The painter has complete manual freedom to explore the canvas and spray at will, with subtle haptic response from the computer delivered through the trigger as the servo restricts it. We used an invisible feedback interface to allow the painter to be immersed in the embodied action of painting instead of relying on a virtual crutch. The painter can select the paint governor aggression modes, from strict restriction to very light restriction (up to none at all), where artists can apply stylizations if they feel confident in their technique.

The Digital Airbrush was first presented in SIGGRAH 2014 [207] as an Emerging Technology demonstration and was very well received. Later we performed a formal user study to examine the effect of computer assistance in the spray painting process, and published the work in Transaction on Graphics [206].

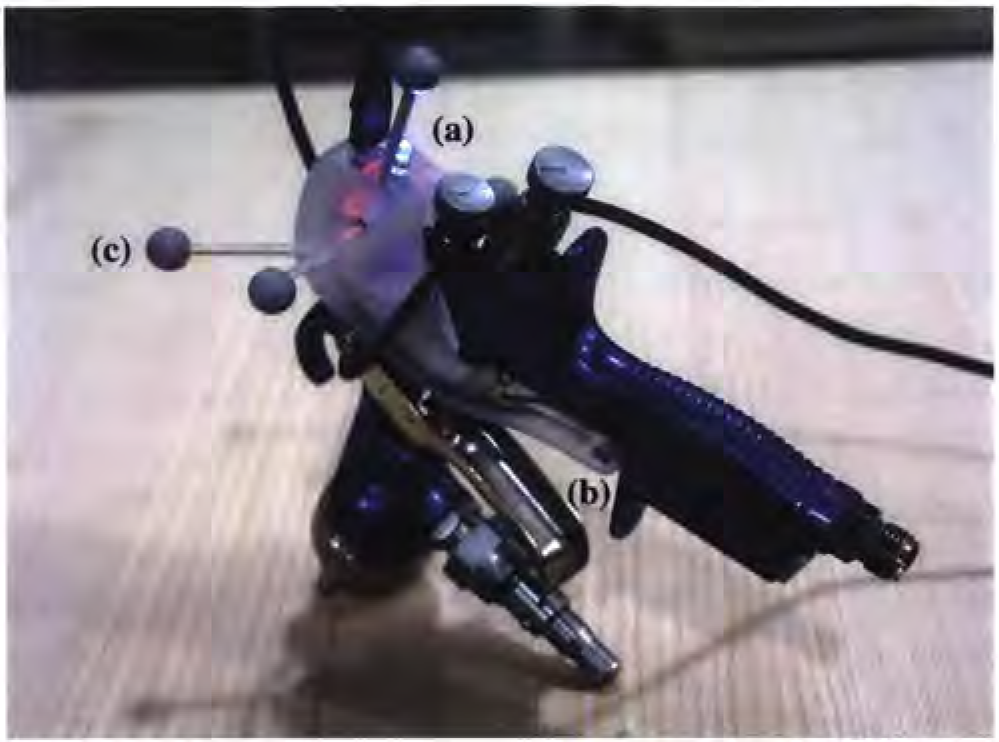


Figure 7-1: The Digital Airbrush for manufacturing. (a) output LED, (b) trigger sensor, (c) tracking markers.

I continued the work on the Digital Airbrush in a manufacturing setting during a collaboration with Steelcase Inc. at Grand Rapids, Michigan in summer 2014. The airbrushes made for manufacturing usage are much more robust than the artistic oriented ones due to the high throughput of paint. Steelcase are using these in the staining stations where they continuously apply coating for pieces of wood as part of the assembly line. The augmentation consisted of an optical tracking system, a magnet-based sensing of the trigger, and output in form of a bright and large LED facing the operator as well as a vibration motor (see figure 7-1 for an illustration). The augmented airbrush was finally chosen to be used as a training tool for factory employees that provides both trainer and trainee with much more information on the staining operation.

7.3 Discussion

In various parts of the thesis (chiefly Chapter 2 and Chapter 4) I touched upon a number of guiding principles rising from the work that can serve as points for discussion around finger augmentation. In essence, finger augmentation leverages, among other things, the following elements: (1) A long and rich tradition in craft, design and symbolism, (2) a unique positioning on the body, and on (3) practiced inherent human behaviors. A more complete list can be found in Section 2.2.4.6. The following discourse combines the outlook from the chapters of this thesis and presents it in a more coherent way, and the next section will flesh these ideas out as directions for future research.

7.3.1 A Rich History

Chapter 2 (Section 2.2.2) provided us with an introduction to the depths of finger-worn fashion and utility. Starting in an unknown point in human existence people started wearing rings and other finger-worn ornaments. The craft of fashioning rings and the art of designing them was preserved through the ages in order to satisfy the many symbolic meanings rings took on: matrimonial bond, protection from evil, religious devotion, and even utility such as a key-ring or a signet-ring for sealing [137; 188]. Augmenting the finger therefore stands on an age-old tradition, which is dual in nature: it suggests that one should be careful about the design as there are "rules" in place, and on the other hand it builds on a socially acceptable form of adornment that is welcome across cultures. This dual nature simultaneously creates opportunity and inhibits it.

To create wearable technology many have leveraged the social acceptability [144; 240] (partial list for the sake of brevity) and other equally important traits of rings, for example: easy to wear and remove, leaves the hands free to move, personal and private, and allow for wearing them for long periods of time. Building on these traits is especially useful for creating inclusive technology, one that can be used by people with a range of needs. Since rings mostly do not carry prejudice or misconception as to their usage, contrasting for example white canes that are very visible and ostracizing, rings can be used for inconspicuous assistance. The same reasons also support the feasibility of ring-like devices as successful commercial products as they can be considered for what they are - objects of fashion.

7.3.2 A Unique Position

Finger worn devices have a doubly unique position: in terms of physical positioning on the body (on the fingers) and a unique position in the universe of wearable computing. Selecting the finger for positioning sensors and actuators is becoming moderately popular with interface researchers (see Chapter 2 for plenty of details on this matter), however the reason for this is likely more complex than straightforward. Fingers are a natural area of sensing for our body in multiple modalities (e.g. touch and vibration, heat and pain), but they are also an incredibly complex collection of muscles, tendons, joints, blood vessels, skin and bones. This led to a large portion of the motor and somatosensory cortices in our brain that is dedicated to operating and sensing the dexterous fingers. One can venture to say that finger augmentation is analogous to natural human evolution, enhancing the intricate array of sensors and actuators we have in our fingertips.

These reasons, physiological and philosophical alike, enticed researchers of wearable user interface to examine the fingers as the natural place for augmentation. Nevertheless, the tight coupling of senses, for example hearing and touch, creates a design and implementation challenge. One of the conclusions of the FingerReader project was that a real-time response from the wearable device is necessary to maintain the illusion of an augmented sense of touch, where touching text will read it out (see Section 4.6). A response that lags or misleads the user has detrimental effect on its utility. Related subjects to this concept, Sensory Substitution and Multi-sensory Integration that were extensively researched over the last 35 years (see for example the body of work of Amir Amedi [172]), do not directly consider augmentation but rehabilitation and neuroplasticity. Finger augmentation presents a singular opportunity to not only combine senses but also enhance them.

7.4 Future Work

While chapter 2 traces the roots of finger augmentation back to the beginning of the human race (albeit in forms of ornaments and jewelry), this research subfield of wearable computing is still in its infancy. Only the last few years have shown a rising interest in developing active finger-worn devices for a variety of uses, where past work was focused on a small set of applications: keyboards, cursor control (computer mouse) and some health signals monitoring. The following are a number of possible future research agendas derived from the current trend and tradition in finger augmentation, as well as my own work.

7.4.1 Assistive Technology

Beyond being a general good motivation for work, assistive technology lends itself rather easily as an application for finger augmentation. Fingers and hands are used by people with impaired senses to communicate and feel (e.g. sign language, braille, listening-by-touch), thus they are already applied regularly towards sensory augmentation, which makes them perfect candidates to do more of the same. Another reason, brought up in the last section, is the high sensitivity of the fingers to certain modalities, which can be complemented with sensors making up for the impaired sense (vision, for example). These reasons, and others mentioned mostly in Chapter 2, open up the design space for assistive finger-wearable technology. The current trend is biased, my work included, towards helping people with an impaired vision [72; 227], however other applications are just as interesting: assisting people with dyslexia, brain trauma, impaired hearing, neurological disorders or elderly people, and crossing over to people in non-permanent disabling situations such as car drivers. Particularly, assistance in actions where the fingers already play a central role (such as reading, writing, or playing a music instrument) can benefit from an augmentation of the fingers.

7.4.1.1 Reading print

Our work in Chapters 4, 5 and 6 made considerable contributions to real-time, assistive, finger-oriented reading of printed material, be it text or music. In spite of this progress, the reading process is still not fluent enough to support real world usage scenarios. One promising direction of future research we uncovered is *Hierarchical Scanning*. In this mode, still in development in our lab, the reader receives hierarchical information on the content of the page before hearing the words or other feedback. Layout information such as: "Page", "Column", "Paragraph", "Heading", "Figure", "Chart" or "Signature box", allows the scanner to then move in and hear more fine-grained information such as individual lines of text, words, graphic information, even down to minuscule printed edge patterns.

A similar idea was postulated by Harvey Lauer, an expert of reading machines for the VI, already in 2003 [110] however Lauer writes about it in a wishful and futuristic way. Lauer suggests that this "closer look" method will give the sought-after feeling of choice and control for a VI reader. Another interesting point Lauer makes is that

VI readers want to participate in the act of reading using technology, instead of *being read to* by technology. We conclude that this future feature will make a great contribution to finger-worn reading systems that support independence.

7.4.2 Natural and Gestural Interaction

A trending subject in Human-Computer Interaction study, natural and gestural interaction is another prime candidate for finger augmentation. The reasons were mentioned in the last section: fingers are already used for gestural language and focusing attention, and they are naturally expressive for their high degree of motoric freedom that is complemented with a dense network of sensors. Work on this topic is already in motion on application such as grasping objects or pointing at them (see [83; 184; 251] and others in Chapter 2), however there is much more to explore: social interaction engaged by touch or gesture, performing arts, self-measurement, learning manual crafts, and other directions.

7.4.3 Communication Devices

Wearable computers came into the communication devices world in waves, from wearable headsets to nowadays' smart watches that replace many functions of a smartphone. This trend is already reaching finger worn devices [145; 169] and is likely to keep growing. In spite of that, not all applications were exhausted and there is still room to examine: in-person communication (e.g. handshakes [69]), communication not with humans (e.g. petting animals, operating appliances), under-utilized modalities in communication (e.g. non-vibrational tactile, thermal), and more. The domain of communication is very likely to be a central pillar in the future of finger augmentation.

7.4.4 Medicine and Physiology

Finger-worn sensors of biological signals, such as heart rate and blood oxygenation (via photoplethysmography), are already in abundance and likely to be integrated into smaller form factors (rings, for example) to provide continuous always-on reading. Beyond that, other biological signals measured with wearable devices such as Electrodermal Activity (EDA) and muscle activity or tension, which correlate with levels of stress, could be integrated into finger-worn form factors leveraging for example on the higher density of sweat glands in the glabrous skin (the skin of the palm). Measuring the pose of the hand via wearable sensors, another trending topic in HCI (e.g. [97; 98]), may be easily tractable with finger-worn devices, or even just a single worn device, and could enrich the offering of input mechanisms for virtual and augmented reality.

7.4.5 Meta and Reflexive Discourse

Finger augmentation, as a new area of research, still needs more defining work to flesh out the social and technical implications. While this thesis offered a definition for the topic, guidelines for design and instances exemplifying these, there is still shortage in work that examines the multi-faceted nature of augmenting fingers. I believe a deeper look into the tradition of finger-worn ornaments and charms could enlighten the technical discussion, and inform the creation of future devices.

7.5 Finger Augmentation Dystopia

Finger augmentation devices (FAD), on all their richness and variety, may not stand the test of time, although most signals point otherwise. We could be simply experiencing a FAD fad these days. Rings on the other hand are more likely to remain. There is therefore a chance that FADs will eventually converge into the ring form factor, since it is the only long-term socially tolerable form. Signs of this are already visible [183]. Another dystopic scenario for FADs is that they disappear in favor of external sensing and actuation of the fingers, which is also on the rise recently (e.g. the Leap Motion), relinquishing the need for actual worn elements. My personal belief is that FADs will ultimately be incorporated in regular-looking rings and worn by people that would naturally wear rings, or start wearing them for their benefit.

It seems these days FADs are in a stage not unlike the early days of the car, when the strict design rules have not been put into place and innovation was in leaps and bounds. Early cars presented very bold design, both mechanically and visually, whereas today the mainstream converged on a somewhat constant class of forms. The same is likely to happen with FADs, where much of the diversity we can see today in form, sensing and actuation will disappear in favor of a practical, perhaps uninteresting, design.

7.6 Thesis Contributions

This thesis makes contributions to the field of finger augmentation that are technical, ethnographic and descriptive. While Chapter 1 discussed these contributions, the following is a brief restatement:

- Computer vision algorithms for finger-perspective imaging
- Evaluation and data-analysis methods for measuring finger-based reading efficacy
- A broad overview of the field of finger augmentation from the last 100 years.

- A design guideline for creating assistive finger augmenting devices, considering aspects in both technical and nontechnical domains.
- Laboratory evaluations of four finger-augmenting devices with qualitative and quantitative methods.

Four finger wearable systems were developed to demonstrate these contributions: The EyeRing, The FingerReader, The Mobile-FingerReader and The MusicReader. Each has a unique operation, design, hardware and software, and were evaluated with different audiences, although they all center around augmentation using a finger-wearable camera. The EyeRing used a wireless stills camera, able to capture one picture at a time, while the other used a wired video camera. Finger-worn cameras are already becoming a crowded design space [263; 227; 257], which is a signal that this vector of development is valid, and "the seeing finger" will soon cease to be a dream.

7.7 Summary

In this thesis I presented an exploration of the space of *Finger Augmentation* for an assisting technology. Finger augmentation finds applications in a wide range of domains, and is steadily growing as a subfield of wearable user interfaces. This document presented four different research projects into the finger-wearable cameras territory, specializing in assistive technology for people with visual impairments. Using the *Assistive Augmentation* principle, these design probes can also be applied towards assisting other target communities. An augmentation of the finger brings together very powerful elements of user interaction: fine motoric ability and heightened sensing second to none in the human body. We believe this combination creates ample opportunity for research and design of user interfaces that tightly integrate the body, mind and computer.

Appendix A

Classification of the Surveyed Finger Augmenting Devices

This appendix includes the complete classification of the works we surveyed in the making of Chapter 2. The abbreviations in the main table are explained below.

In the "type" column:

PT	Patent
A	Academic
PD	Product
C	Concept

In the "position" column:

P	Proximal phalanx
D	Distal phalanx
M	Medial phalanx
W	Whole finger

In the "domain" column:

AC Appliance Control

AS Assistive Technology

BI Biometrics

CM Communication

CR Creative Interaction

GN General

IN Industrial / Robotics

LN Learning

MI Mouse Input

NI Natural Interaction

KI Keyboard Input

VR Virtual or Mixed Reality

SO Social Interaction

In the "form factor" column:

R Ring

D Distal addendum

W Whole finger

N Nail addendum

T Thumb addendum

S Sleeve

P Palm

Domain		AC	CR	MI	KI	AC	KI	MI	KI	BI	KI	MI	MI	CM	AC	BI	MI	IN	KI	MI	CM	MI	AS	KI
Position		P	P	P	P	P	D	M	P	D	D	P	D	P	D		M	P	D	D	M	M	D	M
Form factor		R	R	R	R	R	D	R	R	R	W	R	D	R	N	R	R	R	T	N	R	R	N	R
Wireless						X			X	X		X	X	X		X	X							X
Action	gesture																							
	surface				X		X		X		X		X				X			X			X	
	on device	X		X		X		X		X	X	X		X		X			X		X			
	external		X															X				X		
Output	thermal																							
	radio																							
	magnetic																							
	display																							
	tactile																							
	vibration																						X	
	light																				X	X		
	audio													X										
Input	GPS																							
	magnetic																							
	photo																	X					X	
	button	X				X					X						X		X					
	biometric									X	X					X								
	2D			X								X	X								X			
	proximity					X																		
	pressure						X	X			X	X			X		X			X				
	mechanical																							
	bend																							
	rotation							X																
	temperature											X				X								
	microphone		X											X										
	inertial				X		X		X					X		X								X
Type		PT	PT	PT	A	PT	PT	PT	A	A	A	PT	PT	A	A	PT	PT	PD	A	A	A	PT	A	A
Work		[192]	[149]	[119]	[48]	[17]	[171]	[267]	[49]	[181]	[189]	[44]	[95]	[50]	[132]	[8]	[245]	[232]	[116]	[133]	[144]	[214]	[6]	[108]

Domain		AS	MI	MI	AS	AC	KI	KI	GN	MI	AC	CM	VR	KI	CM	BI	MI	MI	AS	CM	VR	VR	MI	CR
Position		M	W	D	M	P	D	P	P	M	P	P	D	M	P	P	P	M	D	P	D	D	D	P
Form factor		R	W	T	R	W	D	R	S	S	R	R	D	R	R	S	R	R	D	R	D	D	D	R
Wireless				X		X	X		X	X	X	X		X	X	X		X		X				X
Action	gesture					X								X										X
	surface		X				X										X					X		
	on device	X	X	X	X	X			X	X	X	X	X		X		X	X						
	external					X		X				X									X		X	
Output	thermal																							
	radio																							
	magnetic																							
	display						X		X															
	tactile												X						X			X		
	vibration	X			X										X					X	X		X	
	light					X														X			X	
	audio											X												
Input	GPS																							
	magnetic																							
	photo																						X	
	button		X			X	X		X	X	X	X			X		X							
	biometric															X								
	2D			X						X							X							
	proximity																	X						
	pressure			X														X						
	mechanical												X											
	bend					X																		
	rotation																							
	temperature															X								
	microphone											X												
	inertial					X								X		X								X
Type		A	PT	PT	A	A	A	PT	PT	PT	PT	A	A	A	A	A	PT	A	A	A	A	A	A	PD
Work		[70]	[80]	[253]	[5]	[239]	[113]	[35]	[46]	[94]	[150]	[47]	[56]	[100]	[131]	[201]	[14]	[32]	[104]	[107]	[114]	[195]	[263]	[225]

	Domain	MI	BI	AC	NI	VR	VR	AC	AS	GN	CM	GN	AC	GN	AS	SO	LN	AS	VR	VR	BI	CM	BI	BI
	Position	M	P	P	M	W	D	P	P	M	P	P	P	D	P	P	M	D	D	D	D	P	P	P
	Form factor	R	S	R	R	W	D	R	R	R	R	R	R	R	R	R	W	D	D	D	D	R	R	R
	Wireless	X	X	X	X			X	X	X	X	X	X	X	X	X	X				X	X	X	X
Action	gesture			X																				
	surface	X																						
	on device	X				X				X	X		X			X		X	X	X			X	
	external			X	X										X									
Output	thermal																							
	radio																					X		
	magnetic																							
	display											X	X											X
	tactile					X	X					X							X	X				
	vibration								X					X			X	X					X	
	light	X																						
	audio															X								
Input	GPS								X															
	magnetic																							
	photo				X					X														X
	button	X								X						X								
	biometric		X																		X		X	
	2D												X											
	proximity	X			X			X							X									
	pressure	X																X						
	mechanical																							
	bend																							
	rotation																							
	temperature																							
	microphone															X								
	inertial		X	X																	X			
	Type	PT	A	A	A	A	A	C	C	C	C	C	C	C	PT	A	A	A	A	A	A	A	A	C
	Work	[231]	[60]	[112]	[141]	[143]	[197]	[258]	[52]	[148]	[235]	[31]	[242]	[140]	[166]	[23]	[75]	[82]	[99]	[103]	[115]	[208]	[249]	[18]

Domain		SO	SO		VR	MI	KI	MI	AC	VR	LN	GS	VR	VR	VR	VR	NI	AS	CM	AC	CM	VR	CM	AC
Position		P	P	D	D	D	D	P	P	W	P	W	D	D	D	D	D	W	P	P	P	D	P	M
Form factor		R	R		D	D	D	W	R	W	W	W	D	D	D	W	D	W	R	R	R	D	R	S
Wireless		X	X			X		X	X								X	X	X	X	X		X	X
Action	gesture					X	X	X	X															X
	surface	X							X			X		X										
	on device	X			X													X	X	X	X			
	external			X		X											X							
Output	thermal		X																					
	radio																							
	magnetic					X															X			
	display																X		X					
	tactile				X					X			X	X	X	X		X				X		
	vibration			X							X		X						X					X
	light			X														X		X				X
	audio																							X
Input	GPS																							
	magnetic																							
	photo			X														X						
	button																X	X	X	X				
	biometric																							
	2D																							
	proximity	X			X												X							
	pressure								X					X				X						
	mechanical												X			X								
	bend											X												
	rotation																				X			
	temperature																							
	microphone																	X						
	inertial						X	X	X															X
Type		C	C	PT	A	A	A	A	C	A	A	A	A	A	A	A	C	C	PD	PT	A	A	A	A
Work		[69]	[4]	[228]	[7]	[64]	[88]	[163]	[40]	[93]	[102]	[120]	[161]	[170]	[196]	[223]	[36]	[66]	[78]	[213]	[9]	[54]	[57]	[84]

	Domain	LN	MI	AS	MI	MI	VR	BI	MI	NI	MI	CR		MI	GN	MI	MI	MI	KI	NI	MI	NI	BI	GS
	Position	D	P	D	M	P	D	P	D	P	P	P	W	D	D	P	W	P	D	P	D	P	W	M
	Form factor	D	R	D	R	R	D	R	R	R	R	R	W	D	D	R	R	P	D	R	S	R	W	R
	Wireless			X	X	X				X		X				X	X	X				X		X
Action	gesture					X			X	X		X						X			X		X	X
	surface		X			X	X							X										
	on device				X	X					X				X	X	X	X	X			X		
	external			X		X			X					X	X							X		
Output	thermal																							
	radio												X											
	magnetic																							
	display																							
	tactile						X					X												
	vibration	X				X				X														
	light									X				X	X	X								
	audio			X																				
Input	GPS																							
	magnetic																			X	X			
	photo			X										X	X							X		
	button				X	X									X	X		X	X			X		
	biometric																							
	2D				X											X	X							
	proximity									X														
	pressure					X	X				X													
	mechanical																							
	bend							X																
	rotation					X					X													
	temperature									X														
	microphone		X			X																		
	inertial		X						X	X		X						X					X	X
	Type	A	A	C	PD	PT	A	A	A	A	A	A	A	A	C	PD	PT	PT	A	A	A	A	A	A
	Work	[106]	[264]	[111]	[55]	[199]	[34]	[67]	[71]	[96]	[158]	[159]	[244]	[257]	[51]	[191]	[15]	[176]	[12]	[128]	[33]	[68]	[73]	[83]

	Domain	GS	NI	CM	CR	GS	GS	AS	NI	MI	CM	MI	MI	AC	KI	GS	AC	GS	AS	AC	AS	CM	AS	MI
	Position	D	P	D	D	D	D	P		M	P	P	P	P	W	M	W	P	P	D	D	P	D	P
	Form factor	N	R	D	N	D	D	R		R	R	S	S	R	W	R	W	R	R	N	D	R	D	R
	Wireless		X		X			X			X	X	X	X		X	X	X		X				
Action	gesture			X								X	X			X	X							
	surface	X				X	X	X		X			X			X			X				X	X
	on device		X					X				X	X		X		X	X						
	external	X	X					X						X			X		X		X			
Output	thermal																							
	radio				X						X													
	magnetic				X																			
	display			X										X				X						
	tactile															X								
	vibration																	X	X	X	X	X	X	
	light											X					X	X		X		X	X	
	audio																							
Input	GPS																							
	magnetic	X				X																		
	photo		X					X	X				X				X	X	X		X		X	X
	button		X					X				X	X	X	X		X	X	X					
	biometric																							
	2D																							
	proximity																			X				
	pressure																X	X						
	mechanical																							
	bend																							
	rotation																	X						
	temperature																							
	microphone									X								X						
	inertial	X		X		X	X					X	X			X	X							
	Type	A	A	A	A	A	A	A	A	A	C	PD	PD	PD	PT	PT	PT	PT	A	A	A	A	A	A
	Work	[121]	[185]	[229]	[240]	[251]	[252]	[153]	[175]	[146]	[81]	[151]	[236]	[218]	[85]	[105]	[168]	[200]	[203]	[234]	[72]	[169]	[226]	[97]

Domain		AS	BI	BI	GS	GS	KI	CM	CM	MI	MI
Position		P	P	D	P	P	P	P	P	D	W
Form factor		S	R	D	R	R	R	R	R	S	S
Wireless			X	X	X	X	X	X	X	X	
Action	gesture				X	X	X				
	surface	X								X	X
	on device					X	X	X		X	
	external	X		X						X	
Output	thermal										
	radio										
	magnetic										
	display							X			
	tactile										
	vibration	X			X			X	X		
	light				X				X	X	
	audio										
Input	GPS										
	magnetic										
	photo			X						X	X
	button					X		X		X	
	biometric		X								
	2D										
	proximity	X									
	pressure	X					X			X	
	mechanical										
	bend										
	rotation										
	temperature										
	microphone										
	inertial				X	X	X			X	
Type		A	A	PD	PD	PD	PD	PD	PD	PD	PD
Work		[194]	[74]	[3]	[125]	[156]	[182]	[145]	[183]	[174]	[126]

Bibliography

[1] Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV'04*, pages 1–22, 2004.

[2] Wikipedia: Thimble. `http://en.wikipedia.org/wiki/Thimble`, October 2014.

[3] Advanced Realtime Tracking. ART fingertracking, 2014.

[4] Alaska Jewelry. Remember ring, April 2008.

[5] T. Amemiya, J. Yamashita, K. Hirota, and M. Hirose. Virtual leading blocks for the deaf-blind: a real-time way-finder by verbal-nonverbal hybrid interface and high-density RFID tag space. In *Virtual Reality, 2004. Proceedings. IEEE*, pages 165–287, March 2004.

[6] Hideyuki Ando, Takeshi Miki, Masahiko Inami, and Taro Maeda. SmartFinger: Nail-mounted tactile display. In *ACM SIGGRAPH 2002 Conference Abstracts and Applications*, SIGGRAPH '02, pages 78–78, New York, NY, USA, 2002. ACM.

[7] Takafumi Aoki, Hironori Mitake, Danial Keoki, Shoichi Hasegawa, and Makoto Sato. Wearable haptic device to present contact sensation based on cutaneous sensation using thin wire. In *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*, ACE '09, pages 115–122, New York, NY, USA, 2009. ACM.

[8] H.H. Asada, K.W. Chang, K.Y.S. Siu, and B.H. Yang. *Patient monitoring finger ring sensor.* Google Patents, October 1999. US Patent 5,964,701.

[9] Daniel Ashbrook, Patrick Baudisch, and Sean White. Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2043–2046, New York, NY, USA, 2011. ACM.

[10] Tetsuaki Baba, Yuya Kikukawa, Toshiki Yoshiike, Tatsuhiko Suzuki, Rika Shoji, Kumiko Kushiyama, and Makoto Aoki. Gocen: A Handwritten Notational Interface for Musical Performance and Learning Music. In *ACM SIGGRAPH 2012 Emerging Technologies*, SIGGRAPH '12, pages 9:1–9:1, New York, NY, USA, 2012. ACM.

[11] S. Bajaj and B.J. LaVaque. Finger palette, August 23 2012. US Patent App. 13/411,552.

[12] B. Bajer, I.S. MacKenzie, and M. Baljko. Huffman base-4 text entry glove (h4 teg). In *Wearable Computers (ISWC), 2012 16th International Symposium on*, pages 41–47, June 2012.

[13] Javier Bajo, Miguel A Sánchez, Vidal Alonso, Roberto Berjón, Juan A Fraile, and Juan M Corchado. A distributed architecture for facilitating the integration of blind musicians in symphonic orchestras. *Expert Systems with Applications*, 37(12):8508–8515, 2010.

[14] P.M. Baughman. *Wearable ergonomic computer mouse.* Google Patents, February 2005. US Patent 6,850,224.

[15] I. Bell. *A flexible input device worn on a finger.* Google Patents, November 2012. WO Patent App. PCT/GB2011/051,001.

[16] Mohamed Benali-Khoudja, Moustapha Hafez, Jean-Marc Alexandre, and Abderrahmane Kheddar. Tactile interfaces: a state-of-the-art survey. In *Int. Symposium on Robotics*, volume 31. Citeseer, 2004.

[17] E.S. Bennett. Method and apparatus for a weapon firing safety system, October 31 1995. US Patent 5,461,812.

[18] Julien Bergignat and Laurence Charron-Huysecom. UV ring, December 2008.

[19] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. Vizwiz: Nearly real-time answers to visual questions. In *Proc. of UIST*, pages 333–342. ACM, 2010.

[20] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 333–342, New York, NY, USA, 2010. ACM.

[21] Mark Billinghurst. Wearable computers. `http://www.slideshare.net/marknb00/wearable2013`, October 2013.

[22] Alan W Black and Kevin A Lenzo. Flite: a small fast run-time synthesis engine. In *4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis*, 2001.

[23] Robert S Brewer, Samuel RH Joseph, Guanghong Yang, Neil Scott, and Daniel Suthers. SocialSense: A system for social environment awareness. In *UbiComp'08 Workshop W1–Devices that Alter Perception (DAP 2008)*, 2008.

[24] Margaret F Brinig. Rings and Promises. *Journal of Law, Economics, & Organization*, pages 203–215, 1990.

[25] W. Buxton, M. Billinghurst, Y. Guiard, A. Sellen, and S. Zhai. Human input to computer systems: theories, techniques and technology. `http://www.billbuxton.com/inputManuscript.html`, 2011.

[26] Alfredo Capozzi, Roberto De Prisco, Michele Nasti, and Rocco Zaccagnino. Musica Parlata: A Methodology to Teach Music to Blind People. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 245–246, New York, NY, USA, 2012. ACM.

[27] Jeremy I. M. Carpendale and Ailidh B. Carpendale. The development of pointing: From personal directedness to interpersonal direction. *Human development*, 53(3):110–126, 2010.

[28] Ben P Challis and Alistair DN Edwards. Weasel: A system for the Non-visual Presentation of Music Notation. In *Proceedings of ICCHP*, pages 113–120, Karlsruhe, Germany, 2000. Osterreichische Computer Gesellschaf.

[29] Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, Kai-Yin Cheng, Chao-Huai Su, Mike Y. Chen, Wen-Huang Cheng, and Bing-Yu Chen. FingerPad: Private and Subtle Interaction Using Fingertips. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 255–260, New York, NY, USA, 2013. ACM.

[30] Charles Baumann. Finger-guard, May 1916.

[31] Charles Windlin. Digitus - watch ring, October 2007.

[32] Ranajit Chatterjee and F. Matsuno. Design of a touch sensor based single finger operated wearable user-interface terminal. In *SICE-ICASE, 2006. International Joint Conference*, pages 4142–4147, October 2006.

[33] Ke-Yu Chen, Kent Lyons, Sean White, and Shwetak Patel. uTrack: 3d input using two magnetic sensors. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13, pages 237–244, New York, NY, USA, 2013. ACM.

[34] Francesco Chinello, Monica Malvezzi, Claudio Pacchierotti, and Domenico Prattichizzo. A three DoFs wearable tactile display for exploration and manipulation of virtual objects. In *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2012.

[35] J.K. Cho and S. Lee. *Wearable finger montion sensor for sensing finger motion and method of sensing finger motion using the same.* Google Patents, December 2004. US Patent App. 10/876,526.

[36] Choi Hyong-Suk, Jung Ji-hye, and Yoo-Jin Park. My fingers are so smart, they measure!, September 2010.

[37] David Crombie, Sijo Dijkstra, Edmar Schut, and Natasha Lindsay. Spoken Music: Enhancing Access to Music for the Print Disabled. In Klaus Miesenberger, Joachim Klaus, and Wolfgang Zagler, editors, *Computers Helping People with Special Needs*, volume 2398 of *Lecture Notes in Computer Science*, pages 667–674. Springer Berlin Heidelberg, 2002.

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

[39] EE Fournier d'Albe. On a type-reading optophone. *Proceedings of the Royal Society of London. Series A*, 90(619):373–375, 1914.

[40] Denso Corporation. Yubi-nya remote control ring, March 2009.

[41] Paul Dourish. *Where the action is: the foundations of embodied interaction.* MIT press, 2004.

[42] Yasmine N. El-Glaly, Francis Quek, Tonya L. Smith-Jackson, and Gurjot Dhillon. It is Not a Talking Book;: It is More Like Really Reading a Book! In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 277–278, New York, NY, USA, 2012. ACM.

[43] Nobuo Ezaki, Marius Bulacu, and Lambert Schomaker. Text detection from natural scene images: towards a system for visually impaired persons. In *Proc. of ICPR*, volume 2, pages 683–686, 2004.

[44] Lee Felsenstein and Weijia Wang. *Wearable context sensitive user interface for interacting with plurality of electronic devices of interest to the user.* Google Patents, November 1998. US Patent 5,832,296.

[45] Alicia Fornés and Gemma Sánchez. Analysis and Recognition of Music Scores. *Handbook of Document Image Processing and Recognition*, pages 749–774, 2014.

[46] E.C. Frederick, K. Marcucelli, and C. Wojcieszak. *Housing for electronic device wearable on user's finger.* Google Patents, September 2004. EP Patent App. EP20,030,024,857.

[47] Masaaki Fukumoto. A finger-ring shaped wearable HANDset based on bone-conduction. In *International Symposium on Wearable Computers*, pages 10–13, 2005.

[48] Masaaki Fukumoto and Yasuhito Suenaga. FingeRing: A full-time wearable interface. In *Conference Companion on Human Factors in Computing Systems*, CHI '94, pages 81–82, New York, NY, USA, 1994. ACM.

[49] Masaaki Fukumoto and Yoshinobu Tonomura. Body coupled FingerRing: Wireless wearable keyboard. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 147–154, New York, NY, USA, 1997. ACM.

[50] Masaaki Fukumoto and Yoshinobu Tonomura. Whisper: A wristwatch style wearable handset. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 112–119, New York, NY, USA, 1999. ACM.

[51] Yoshimasa Furuyama. Ubi-camera. `http://www.gizmag.com/ubi-camera-fingers-viewfinder/21997/`, March 2012.

[52] Gail Knight. GPS ring, July 2007.

[53] John Galeotti, Samantha Horvath, Roberta Klatzky, Brock Nichol, Mel Siegel, and George Stetten. FingerSight^{TM}: fingertip control and haptic sensing of the visual environment. In *ACM SIGGRAPH 2008 new tech demos*, page 16. ACM, 2008.

[54] Paulo Gallotti, Alberto Raposo, and Luciano Soares. v-glove: A 3d virtual touch interface. In *Symposium on Virtual Reality*, 2011.

[55] Genius. Thumb controller, 2011.

[56] Florian Gosselin, Tanguy Jouan, Julien Brisset, and Claude Andriot. Design of a wearable haptic interface for precise finger interactions in large virtual environments. In *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 202–207, 2005.

[57] H. Goto and H. Iwasaki. A low profile wideband monopole antenna with double finger ring for BAN. In *Antennas and Propagation in Wireless Communications (APWC), 2011 IEEE-APS Topical Conference on*, pages 409–412, September 2011.

[58] Toshiyuki Gotoh, Reiko Minamikawa-Tachino, and Naoyoshi Tamura. A Web-based Braille Translation for Digital Music Scores. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, pages 259–260, New York, NY, USA, 2008. ACM.

[59] T.S. Grover. Basketball shooting training device and method, November 8 2012. US Patent App. 13/465,861.

[60] Hyonyoung Han, Min-Joon Kim, and Jung Kim. Development of real-time motion artifact reduction algorithm for a wearable photoplethysmography. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 1538–1541, August 2007.

[61] Shehzad Muhammad Hanif and Lionel Prevost. Texture based text detection in natural scene images-a help to blind and visually impaired persons. In *CVHI*, 2007.

[62] Susumu Harada, Daisuke Sato, Dustin W. Adams, Sri Kurniawan, Hironobu Takagi, and Chieko Asakawa. Accessible photo album: Enhancing the photo sharing experience for people with visual impairment. In *Proc. of CHI*, pages 2127–2136. ACM, 2013.

[63] A.H. Harris. *Hand-switch for automobiles.* Google Patents, October 1918. US Patent 1,280,457.

[64] Chris Harrison and Scott E. Hudson. Abracadabra: Wireless, high-precision, and unpowered finger input for very small mobile devices. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, UIST '09, pages 121–124, New York, NY, USA, 2009. ACM.

[65] Steve Harrison, Phoebe Sengers, and Deborah Tatar. Making epistemological trouble: Third-paradigm hci as successor science. *Interacting with Computers*, 23(5):385–392, 2011.

[66] Erik Hedberg and Zack Bennett. Thimble - There's a Thing for That, December 2010.

[67] Pilwon Heo and Jung Kim. Finger flexion force sensor based on volar displacement of flexor tendon. In *International Conference on Robotics and Automation*, pages 1392–1397, 2012.

[68] Anuruddha Hettiarachchi, Suranga Nanayakkara, Kian Peen Yeo, Roy Shilkrot, and Pattie Maes. FingerDraw: more than a digital paintbrush. In *Proceedings of the 4th Augmented Human International Conference*, pages 1–4. ACM, 2013.

[69] Hideaki Matsui. Information ring, January 2008.

[70] Michitaka Hirose and Tomohiro Amemiya. Wearable finger-braille interface for navigation of deaf-blind in ubiquitous barrier-free space. In *Proceedings of the HCI International*, volume 4, pages 1417–1421, 2003.

[71] Tatsuya Horie, Tsutomu Terada, Takuya Katayama, and Masahiko Tsukamoto. A pointing method using accelerometers for graphical user interfaces. In *Proceedings of the 3rd Augmented Human International Conference*, AH '12, pages 12:1–12:8, New York, NY, USA, 2012. ACM.

[72] S. Horvath, J. Galeotti, Bing Wu, R. Klatzky, M. Siegel, and G. Stetten. FingerSight: Fingertip haptic sensing of the visual environment. *Translational Engineering in Health and Medicine, IEEE Journal of*, 2:1–9, December 2014.

[73] Christopher-Eyk Hrabia, Katrin Wolf, and Mathias Wilhelm. Whole hand modeling using 8 wearable sensors: Biomechanics for hand pose prediction. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, pages 21–28, New York, NY, USA, 2013. ACM.

[74] Cheng-Yang Huang, Ming-Che Chan, Chien-Yue Chen, and Bor-Shyh Lin. Novel wearable and wireless ring-type pulse oximeter with multi-detectors. *Sensors*, 14(9):17586–17599, 2014.

[75] Kevin Huang, Ellen Yi-Luen Do, and Thad Starner. PianoTouch: A wearable haptic piano instruction system for passive learning of piano skills. In *International Symposium on Wearable Computers*, pages 41–44, 2008.

[76] Jochen Huber, Jun Rekimoto, Masahiko Inami, Roy Shilkrot, Pattie Maes, Wong Meng Ee, Graham Pullin, and Suranga Chandima Nanayakkara. Workshop on assistive augmentation. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 103–106, New York, NY, USA, 2014. ACM.

[77] Sungjae Hwang, Dongchul Kim, Sang-won Leigh, and Kwang-yun Wohn. NailSense: fingertip force as a new input modality. In *Proceedings of the adjunct publication of the 26th annual ACM symposium on User interface software and technology*, pages 63–64. ACM, 2013.

[78] Hybratech. ORB ring, 2010.

[79] Virginia A Jacko, Jin Ho Choi, Audrey Carballo, Brian Charlson, and J Elton Moore. A New Synthesis of Sound and Tactile Music Code Instruction in a Pilot Online Braille Music Curriculum. *Journal of Visual Impairment & Blindness (Online)*, 109(2):153, 2015.

[80] E.D. Jarra. *Finger securable computer input device.* Google Patents, July 2003. US Patent 6,587,090.

[81] Jeff Fransen, Roy Taragan, and Netanel Biton. Ring tuit, September 2013.

[82] Li Jiang, Mark R Cutkosky, Juhani Ruutiainen, and Roope Raisamo. Improving finger force control with vibrational haptic feedback for multiple sclerosis. In *Proceedings of the IASTED International Conference on Telehealth/Assistive Technologies*, pages 110–115. ACTA Press, 2008.

[83] Lei Jing, Zixue Cheng, Yinghui Zhou, Junbo Wang, and Tongjun Huang. Magic ring: A self-contained gesture input device on finger. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, MUM '13, pages 39:1–39:4, New York, NY, USA, 2013. ACM.

[84] Lei Jing, Kaoru Yamagishi, Junbo Wang, Yinghui Zhou, Tongjun Huang, and Zixue Cheng. A unified method for multiple home appliances control through static finger gestures. In *Symposium on Applications and the Internet*, 2011.

[85] K.T. Johnson. *Data input device.* Google Patents, December 2013. US Patent 8,605,041.

[86] Matt Jones and Gary Marsden. *Mobile interaction design.* John Wiley & Sons, 2006.

[87] Joseph R Zazzara. *Writing implement holder.* Google Patents, August 1971. US Patent 3,596,964.

[88] Yuki Kanai, Makoto Oka, and Hirohiko Mori. *Manipulation with Fingers in a 3-D Physical Space.* 2009.

[89] Shaun K. Kane, Brian Frey, and Jacob O. Wobbrock. Access Lens: A Gesture-based Screen Reader for Real-world Documents. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 347–350, New York, NY, USA, 2013. ACM.

[90] Shaun K. Kane, Chandrika Jayant, Jacob O. Wobbrock, and Richard E. Ladner. Freedom to Roam: A Study of Mobile Device Adoption and Accessibility for People with Visual and Motor Disabilities. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '09, pages 115–122, New York, NY, USA, 2009. ACM.

[91] Hsin-Liu (Cindy) Kao, Artem Dementyev, Joseph A. Paradiso, and Chris Schmandt. NailO: Fingernails As an Input Surface. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3015–3018, New York, NY, USA, 2015. ACM.

[92] Kawai Musical Instruments Manufacturing Co., Ltd. Gakufu, August 2011.

[93] H. Kawasaki, Y. Doi, S. Koide, T. Endo, and T. Mouri. Hand haptic interface incorporating 1d finger pad and 3d fingertip force display devices. In *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, pages 1869–1874, July 2010.

[94] S. Kenin. *Index-finger computer mouse.* Google Patents, April 2004. US Patent App. 10/279,800.

[95] R.H. Kent and J.D. Wentz. *Finger operated digital input device.* Google Patents, January 1998. US Patent 5,706,026.

[96] H. Ketabdar, P. Moghadam, and M. Roshandel. Pingu: A new miniature wearable device for ubiquitous computing environments. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on*, pages 502–506, July 2012.

[97] Wolf Kienzle and Ken Hinckley. LightRing: Always-available 2d input on any surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 157–160, New York, NY, USA, 2014. ACM.

[98] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 167–176, New York, NY, USA, 2012. ACM.

[99] Seung-Chan Kim, Chong Hui Kim, Tae-Hon Yang, Gi-Hun Yang, Sung-Chul Kang, and Dong-Soo Kwon. SaLT: Small and lightweight tactile display using ultrasonic actuators. In *IEEE International Symposium on Robot and Human Interactive Communication*, 2008.

[100] Yoon Sang Kim, Byung Seok Soh, and Sang-Goog Lee. A new wearable input device: SCURRY. *Industrial Electronics, IEEE Transactions on*, 52(6):1490–1499, December 2005.

[101] C. Knowles. *Self-defense ring.* Google Patents, September 2005. US Patent App. 11/061,177.

[102] Daniel Kohlsdorf and Thad Starner. Mobile music touch: The effect of primary tasks on passively learning piano sequences. In *International Symposium on Wearable Computers*, pages 1–8, 2010.

[103] Ig Mo Koo, Kwangmok Jung, Ja-Choon Koo, Jae-Do Nam, Young Kwan Lee, and Hyouk-Ryeol Choi. Development of soft-actuator-based wearable tactile display. *Robotics, IEEE Transactions on*, 24(3):549–558, June 2008.

[104] Igmo Koo, Kwangmok Jung, Jachoon Koo, Jea-do Nam, Youngkwan Lee, and Hyoukryeol Choi. Wearable tactile display based on soft actuator. In *International Conference on Robotics and Automation*, pages 2220–2225, 2006.

[105] B. Kruse and R. Steger. *Method and apparatus for wearable remote interface device.* Google Patents, March 2013. US Patent 8,405,612.

[106] Yuichi Kurita, Minoru Shinohara, and Jun Ueda. Wearable sensorimotor enhancer for a fingertip based on stochastic resonance. In *International Conference on Robotics and Automation*, pages 3790–3795, 2011.

[107] Jean-Baptiste Labrune and Wendy Mackay. Telebeads: Social network mnemonics for teenagers. In *Proceedings of the 2006 Conference on Interaction Design and Children*, IDC '06, pages 57–64, New York, NY, USA, 2006. ACM.

[108] A.H.F. Lam and W.J. Li. MIDS: GUI and TUI in mid-air using MEP sensors. In *Control and Automation, 2002. ICCA. Final Program and Book of Abstracts. The 2002 International Conference on*, pages 86–87, June 2002.

[109] J.A. Lambert. *Hand mounted aviation night vision illuminating device.* Google Patents, June 1992. US Patent 5,124,892.

[110] Harvey Lauer. The reading machine that hasn't been built yet. *AccessWorld*, 4(2), March 2003.

[111] Hansub Lee. Finger Reader, September 2011.

[112] Jupyung Lee, Seung-Ho Lim, Jong-woon Yoo, Ki-woong Park, Hyun-jin Choi, and Kyu Ho Park. A ubiquitous fashionable computer with an i-throw device on a location-based service environment. In *Advanced Information Networking and Applications*, pages 59–65, 2007.

[113] Seongil Lee and SangHyuk Hong. Design of an integrated wearable multimedia interface for in-vehicle telematics. In Kiyoharu Aizawa, Yuichi Nakamura, and Shin'ichi Satoh, editors, *Advances in Multimedia Information Processing - PCM 2004*, volume 3332 of *Lecture Notes in Computer Science*, pages 113–120. Springer Berlin Heidelberg, 2004.

[114] Shang Ping Lee, Adrian David Cheok, James Keng Soon Teh, Goh Pae Lyn Debra, Chio Wen Jie, Wang Chuang, and Farzam Farbiz. A mobile pet wearable computer and mixed reality system for human-poultry interaction through the internet. *Personal and Ubiquitous Computing*, 10(5):301–317, 2006.

[115] Young-Dong Lee, Sang-Joong Jung, Yong-Su Seo, and Wan-Young Chung. Measurement of motion activity during ambulatory using pulse oximeter and triaxial accelerometer. 2008.

[116] Juha Lehikoinen and Mika Röykkee. N-fingers: a finger-based interaction technique for wearable computers. *Interacting with Computers*, 13(5):601–625, 2001.

[117] Vladimir I Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, volume 10, page 707, 1966.

[118] Vincent Lévesque. Blindness, technology and haptics. *Center for Intelligent Machines*, pages 19–21, 2005.

[119] N.A. Levine. *Finger worn graphic interface device.* Google Patents, September 1990. US Patent 4,954,817.

[120] Kang Li, I.-Ming Chen, and Song Huat Yeo. Design and validation of a multi-finger sensing device based on optical linear encoder. In *International Conference on Robotics and Automation*, pages 3629–3634, 2010.

[121] Rong-Hao Liang. Augmenting the input space of portable displays using add-on hall-sensor grid. In *Proceedings of the Adjunct Publication of the 26th Annual ACM Symposium on User Interface Software and Technology*, UIST '13 Adjunct, pages 33–36, New York, NY, USA, 2013. ACM.

[122] Nicola Liberati. Augmented reality and ubiquitous computing: the hidden potentialities of augmented reality. *AI & SOCIETY*, pages 1–12.

[123] John G Linvill and James Charles Bliss. A direct translation reading aid for the blind. *Proc. of the IEEE*, 54(1):40–51, 1966.

[124] Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, and Max Mühlhäuser. Earput: Augmenting behind-the-ear devices for ear-based interaction. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1323–1328, New York, NY, USA, 2013. ACM.

[125] Logbar Inc. Ring: Shortcut everything, 2014.

[126] Logisys. BLACK OPTICAL FINGER MOUSE. Model# MS601BK.

[127] Nawapon Luangnapa, Thongchai Silpavarangkura, Chakarida Nukoolkit, and Pornchai Mongkolnam. Optical Music Recognition on Android Platform. In Borworn Papasratorn, Nipon Charoenkitkarn, Kittichai Lavangnananda, Wichian Chutimaskul, and Vajirasak Vanijja, editors, *Advances in Information Technology*, volume 344 of *Communications in Computer and Information Science*, pages 106–115. Springer Berlin Heidelberg, 2012.

[128] Takuya Maekawa, Yasue Kishino, Yutaka Yanagisawa, and Yasushi Sakurai. Recognizing handheld electrical device usage with hand-worn coil of wire. In Judy Kay, Paul Lukowicz, Hideyuki Tokuda, Patrick Olivier, and Antonio

Krüger, editors, *Pervasive Computing*, volume 7319 of *Lecture Notes in Computer Science*, pages 234–252. Springer Berlin Heidelberg, 2012.

[129] Shachar Maidenbaum, Shlomi Hanassy, Sami Abboud, Galit Buchs, Daniel-Robert Chebat, Shelly Levy-Tzedek, and Amir Amedi. The" eyecane", a new electronic travel aid for the blind: Technology, behavior & swift learning. *Restorative neurology and neuroscience*, 32(6):813–824, 2014.

[130] Steve Mann and Hal Niedzviecki. *Cyborg: Digital Destiny and Human Possibility in the Age of the Wearable Computer.* Doubleday Canada, 2001.

[131] Stefan Marti and Chris Schmandt. Giving the caller the finger: Collaborative responsibility for cellphone interruptions. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1633–1636, New York, NY, USA, 2005. ACM.

[132] Stephen A. Mascaro and H. Harry Asada. Virtual switch human-machine interface using fingernail touch sensors. In *International Conference on Robotics and Automation*, volume 4, pages 2533–2538, 1999.

[133] Stephen A. Mascaro and Haruhiko Asada. Finger posture and shear force measurement using fingernail sensors: Initial experimentation. In *International Conference on Robotics and Automation*, volume 2, pages 1857–1862, 2001.

[134] Yasuhiro Matsuda and Tsuneshi Isomura. Novel mounts of accelerometers of finger braille recognition system. *Engineering Letters*, 20(3), 2012.

[135] M.A. Mattar, A.R. Hanson, and E.G. Learned-Miller. Sign classification using local and meta-features. In *CVPR - Workshops*, pages 26–26. IEEE, June 2005.

[136] Danielle Matthews, Tanya Behne, Elena Lieven, and Michael Tomasello. Origins of the human pointing gesture: a training study. *Developmental Science*, 15(6):817–829, 2012.

[137] James Remington McCarthy. *Rings Through the Ages: An Informal History.* Harper & Brothers, 1945.

[138] D McNeill. Catchments and context: Non-modular factors in speech and gesture. In *Language and Gesture*, chapter 15, pages 312–328. Cambridge University Press, New York, NY, USA, 2000.

[139] David McNeill. *Language and gesture*, volume 2. Cambridge University Press, 2000.

[140] Meng Fandi. Alarming ring, July 2007.

[141] David Merrill and Pattie Maes. Augmenting looking, pointing and reaching gestures to enhance the searching and browsing of physical objects. In *Pervasive Computing*, pages 1–18. Springer, 2007.

[142] David Merrill and Pattie Maes. Augmenting looking, pointing and reaching gestures to enhance the searching and browsing of physical objects. In *Proceedings of the PERVASIVE'07*, pages 1–18. Springer-Verlag, 2007.

[143] Kouta Minamizawa, Souichiro Fukamachi, Hiroyuki Kajimoto, Naoki Kawakami, and Susumu Tachi. Gravity grabber: Wearable haptic display to present virtual mass sensation. In *ACM SIGGRAPH 2007 Emerging Technologies*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[144] Cameron S. Miner, Denise M. Chan, and Christopher Campbell. Digital jewelry: Wearable technology for everyday life. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, pages 45–46, New York, NY, USA, 2001. ACM.

[145] Mota. Mota smart ring, September 2014.

[146] Adiyan Mujibiya, Xiang Cao, Desney S. Tan, Dan Morris, Shwetak N. Patel, and Jun Rekimoto. The sound of touch: On-body touch and gesture sensing based on transdermal ultrasound propagation. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '13, pages 189–198, New York, NY, USA, 2013. ACM.

[147] G. Munari. *Lighting device with resilient fastener for attaching to human finger.* Google Patents, January 2007. US Patent 7,165,859.

[148] David Münscher. Every stalker's dream: Camera ring. `http://www.yankodesign.com/2007/10/22/every-stalkers-dream/`, October 2007.

[149] J. Murad. *Finger ring microphone.* Google Patents, March 1979. US Patent 4,146,757.

[150] W.P. Murray. *TV mute finger ring.* Google Patents, August 2004. US Patent 6,778,380.

[151] Mycestro. Mycestro, 2013.

[152] Suranga Nanayakkara, Roy Shilkrot, and Pattie Maes. Eyering: a finger-worn assistant. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1961–1966. ACM, 2012.

[153] Suranga Nanayakkara, Roy Shilkrot, Kian Peen Yeo, and Pattie Maes. EyeRing: A Finger-worn Input Device for Seamless Interactions with Our Surroundings. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, pages 13–20, New York, NY, USA, 2013. ACM.

[154] NEC. Imaging and recognition solutions. `http://www.nec.co.jp/techrep/en/journal/g11/n03/110302.html`.

[155] E.R. Newton. *Hand signal device.* Google Patents, September 1916. US Patent 1,197,652.

[156] Nod, Inc. Nod, 2014.

[157] Nonin Medical Inc. Finger pulse oximiter, 2015.

[158] Masa Ogata, Yuta Sugiura, Hirotaka Osawa, and Michita Imai. iRing: Intelligent ring using infrared reflection. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 131–136, New York, NY, USA, 2012. ACM.

[159] Masa Ogata, Yuta Sugiura, Hirotaka Osawa, and Michita Imai. Pygmy: A ring-shaped robotic device that promotes the presence of an agent on human hand. In *Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction*, APCHI '12, pages 85–92, New York, NY, USA, 2012. ACM.

[160] Uran Oh and Leah Findlater. Design of and Subjective Response to On-body Input for People with Visual Impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, pages 115–122, New York, NY, USA, 2014. ACM.

[161] Tatsuya Ooka and Kinya Fujita. Virtual object manipulation system with substitutive display of tangential force and slip by control of vibrotactile phantom sensation. In *Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2010.

[162] OrCam. Orcam. `http://www.orcam.com/`, 2014.

[163] Anala Pandit, D. Dand, S. Mehta, S. Sabesan, and A. Daftery. A simple wearable hand gesture recognition device using iMEMS. In *International Conference of Soft Computing and Pattern Recognition*, 2009.

[164] Sue Pavey, Andrew Dodgson, Graeme Douglas, and Ben Clements. Travel, transport, and mobility of people who are blind and partially sighted in the uk. Royal National Institute for the Blind (RNIB), April 2009.

[165] Marcin Pazio, Maciej Niedzwiecki, Ryszard Kowalik, and Jacek Lebiedz. Text detection system for the blind. In *15th european signal processing conference EUSIPCO*, pages 272–276, 2007.

[166] D. Perlman. *Awareness enhancement and monitoring devices for the treatment of certain impulse control disorders.* Google Patents, April 2007. US Patent App. 11/233,972.

[167] J-P Peters, Celine Thillou, and Silvio Ferreira. Embedded reading device for blind people: a user-centered design. In *Proc. of ISIT*, pages 217–222. IEEE, 2004.

[168] A. Poznansky, A. Yahalom, and C. Haim. *Electronic Remote Control Thimble.* Google Patents, October 2013. US Patent App. 13/864,139.

[169] Gilang Andi Pradana, Adrian David Cheok, Masahiko Inami, Jordan Tewell, and Yongsoon Choi. Emotional priming of mobile text messages with ring-shaped wearable device using color lighting and tactile expressions. In *Proceedings of the 5th Augmented Human International Conference*, AH '14, pages 14:1–14:8, New York, NY, USA, 2014. ACM.

[170] Domenico Prattichizzo, Francesco Chinello, Claudio Pacchierotti, and Kouta Minamizawa. RemoTouch: A system for remote touch experience. In *IEEE International Symposium on Robot and Human Interactive Communication*, pages 676–679, 2010.

[171] K.R. Prince. *Method of entering information into a computer.* Google Patents, December 1996. US Patent 5,581,484.

[172] Michael J Proulx, Maurice Ptito, and Amir Amedi. Multisensory integration, sensory substitution and visual rehabilitation. *Neuroscience & Biobehavioral Reviews*, 41:1–2, 2014.

[173] G. Pullin. *Design meets disability.* MIT Press, 2009.

[174] Rami Parham, Yuval Ben-Zeev, Jonatan Ary, Danny Lazar, and Gil Rotgold. Muv bird, May 2014.

[175] Shanaka Ransiri and Suranga Nanayakkara. SmartFinger: An augmented finger as a seamless 'channel' between digital and physical objects. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, pages 5–8, New York, NY, USA, 2013. ACM.

[176] C.S. Ranta, S.N. Bathiche, and M. Chatterjee. *Wearable computer pointing device.* Google Patents, February 2012. US Patent 8,125,448.

[177] Ana Rebelo, Ichiro Fujinaga, Filipe Paszkiewicz, AndreR.S. Marcal, Carlos Guedes, and JaimeS. Cardoso. Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1(3):173–190, 2012.

[178] E. Reese and A. Cox. Quality of adult book reading affects children's emergent literacy. *Developmental Psychology*, 35(1):20–28, 1999.

[179] Jun Rekimoto. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Proceedings of the ISWC '01*, pages 21–27. IEEE Computer Society, 2001.

[180] Jun Rekimoto and Katashi Nagao. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of the 8th annual ACM symposium on User interface and software technology*, pages 29–36. ACM, 1995.

[181] Sokwoo Rhee, Boo-Ho Yang, Kuowei Chang, and H.H. Asada. The ring sensor: a new ambulatory wearable sensor for twenty-four hour patient monitoring. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 4, pages 1906–1909 vol.4, October 1998.

[182] RHLvision Technologies Pvt.Ltd. fin, 2014.

[183] Ringly. Ringly. `https://ringly.com/`, 2014.

[184] Mikko J. Rissanen, Owen Noel Newton Fernando, Horathalge Iroshan, Samantha Vu, Natalie Pang, and Schubert Foo. Ubiquitous Shortcuts: Mnemonics by Just Taking Photos. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1641–1646, New York, NY, USA, 2013. ACM.

[185] MikkoJ. Rissanen, Samantha Vu, OwenNoelNewton Fernando, Natalie Pang, and Schubert Foo. Subtle, natural and socially acceptable interaction techniques for ringterfaces â finger-ring shaped user interfaces. In Norbert Streitz and Constantine Stephanidis, editors, *Distributed, Ambient, and Pervasive Interactions*, volume 8028 of *Lecture Notes in Computer Science*, pages 52–61. Springer Berlin Heidelberg, 2013.

[186] JW Roach and JE Tatem. Using domain knowledge in low-level visual processing to interpret handwritten music: an experiment. *Pattern recognition*, 21(1):33–44, 1988.

[187] Yvonne Rogers. Hci theory: classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics*, 5(2):1–129, 2012.

[188] Elizabeth J Roop. *History and Meaning of Symbolic Rings.* PhD thesis, Ball State University, May 2011.

[189] Robert Rosenberg. *Computing without mice and keyboards: text and graphic input devices for mobile computing.* PhD thesis, University College London, 1998.

[190] Kimiko Ryokai, Stefan Marti, and Hiroshi Ishii. Designing the world as your palette. In *Proceesings of CHI '05*, pages 1037–1049, New York, NY, USA, 2005. ACM.

[191] Saar Shai and Efrat Barit. Ringbow, 2012.

[192] S. Samuel. Finger mount assembly for control switch, December 28 1965. US Patent 3,226,501.

[193] Andreas Sand, Christian NS Pedersen, Thomas Mailund, and Asbjørn Tølbøl Brask. Hmmlib: A c++ library for general hidden markov models exploiting modern cpus. In *2010 Ninth International Workshop on Parallel and Distributed Methods in Verification/2010 Second International Workshop on High Performance Computational Systems Biology*, pages 126–134. IEEE, 2010.

[194] Santiago Alfaro. Digital synesthesia, May 2014.

[195] I. Sarakoglou, M. Bezdicek, N. Tsagarakis, and D.G. Caldwell. Free to touch: A portable tactile display for 3d surface texture exploration. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 3587–3592, October 2006.

[196] S. Scheggi, G. Salvietti, and D. Prattichizzo. Shape and weight rendering for haptic augmented reality. In *RO-MAN, 2010 IEEE*, pages 44–49, September 2010.

[197] Robert Scheibe, Mathias Moehring, and Bernd Froehlich. Tactile feedback at the finger tips for improved direct interaction in immersive environments. In *3D User Interfaces, 2007. 3DUI'07. IEEE Symposium on.* IEEE, 2007.

[198] Andrew Sears and Vicki Hanson. Representing users in accessibility research. In *Proc. of CHI*, pages 2235–2238. ACM, 2011.

[199] S. Shai. *Finger-worn device and interaction methods and communication methods.* Google Patents, September 2011. US Patent App. 13/049,925.

[200] S. Shai. *Finger-worn devices and related methods of use.* Google Patents, May 2013. EP Patent App. EP20,130,152,219.

[201] P. Shaltis, L. Wood, A. Reisner, and H. Asada. Novel design for a wearable, rapidly deployable, wireless noninvasive triage sensor. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 3567–3570, January 2005.

[202] Huiying Shen and James M Coughlan. Towards a real-time system for finding and reading signs for visually impaired users. In *Computers Helping People with Special Needs*, pages 41–47. Springer, 2012.

[203] Roy Shilkrot, Jochen Huber, Connie Liu, Pattie Maes, and Suranga Chandima Nanayakkara. FingerReader: A wearable device to support text reading on the go. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 2359–2364, New York, NY, USA, 2014. ACM.

[204] Roy Shilkrot, Jochen Huber, Connie Liu, Pattie Maes, and Suranga Chandima Nanayakkara. Fingerreader: A wearable device to support text reading on the go. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2359–2364. ACM, 2014.

[205] Roy Shilkrot, Jochen Huber, Wong Meng Ee, Pattie Maes, and Suranga Chandima Nanayakkara. FingerReader: A Wearable Device to Explore Printed Text on the Go. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2363–2372, New York, NY, USA, 2015. ACM.

[206] Roy Shilkrot, Pattie Maes, Joseph A. Paradiso, and Amit Zoran. Augmented airbrush for computer aided painting (cap). *ACM Trans. Graph.*, 34(2):19:1–19:11, March 2015.

[207] Roy Shilkrot, Pattie Maes, and Amit Zoran. Physical rendering with a digital airbrush. In *ACM SIGGRAPH 2014 Emerging Technologies*, SIGGRAPH '14. ACM, 2014.

[208] Yuichi Shimizu and Nobuhiro Kuga. A finger-ring antenna for BAN and PAN applications. *ISAP2008, Taipei, Oct*, 2008.

[209] K. Shinohara and J. Tenenberg. A blind person's interactions with technology. *Communications of the ACM*, 52(8):58–66, 2009.

[210] K. Shinohara and J.O. Wobbrock. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 705–714. ACM, 2011.

[211] Kristen Shinohara and Josh Tenenberg. A blind person's interactions with technology. *Commun. ACM*, 52(8):58–66, August 2009.

[212] Kristen Shinohara and Jacob O. Wobbrock. In the shadow of misperception: Assistive technology use and social interactions. In *Proc. of CHI*, pages 705–714. ACM, 2011.

[213] C.O.B. Shirey. *Hands free light emitting diode (LED) illumination device.* Google Patents, April 2010. US Patent 7,703,937.

[214] J.L. Sibert and M. Gokturk. *Direct pointing apparatus and method therefor.* Google Patents, February 2001. US Patent 6,184,863.

[215] Itiro Siio, Toshiyuki Masui, and Kentaro Fukuchi. Real-world interaction using the fieldmouse. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, UIST '99, pages 113–119, New York, NY, USA, 1999. ACM.

[216] Siri. `www.apple.com/ios/siri`.

[217] Mary A. Smaligo. Resources for Helping Blind Music Students. *Music Educators Journal*, 85(2):pp. 23–26+45, 1998.

[218] SmartyRing. SmartyRing. `http://smartyring.com`, 2013.

[219] O. Smith. *Ring pen device.* Google Patents, June 2005. US Patent 6,910,821.

[220] Ray Smith. An overview of the tesseract OCR engine. In *Proc. of ICDAR*, volume 2, pages 629–633. IEEE, 2007.

[221] SnapNPlay music. SnapNPlay music, May 2013.

[222] Byung Seok Soh, Yoon Sang Kim, and Sang-Goog Lee. Improving the usability of a wearable input device SCURRYTM. 2004.

[223] M. Solazzi, A. Frisoli, and M. Bergamasco. Design of a novel finger haptic interface for contact and orientation display. In *Haptics Symposium, 2010 IEEE*, pages 129–132, March 2010.

[224] Thanachai Soontornwutikul, Nutcha Thananart, Aphinun Wantanareeyachart, Chakarida Nukoolkit, and Chonlameth Arpnikanondt. Optical Music Recognition on Windows Phone 7. In Phayung Meesad, Herwig Unger, and Sirapat Boonkrong, editors, *The 9th International Conference on Computing and InformationTechnology (IC2IT2013)*, volume 209 of *Advances in Intelligent Systems and Computing*, pages 239–248. Springer Berlin Heidelberg, 2013.

[225] SourceAudio. Hot hand, January 2006.

[226] Lee Stearns, Ruofei Du, Uran Oh, Yumeng Wang, Leah Findlater, Rama Chellappa, and Jon E Froehlich. The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind. September 2014.

[227] Lee Stearns, Ruofei Du, Uran Oh, Yumeng Wang, Leah Findlater, Rama Chellappa, and Jon E Froehlich. The design and preliminary evaluation of a finger-mounted camera and feedback system to enable reading of printed text for the blind. In *Workshop on Assistive Computer Vision and Robotics, ECCV*. Springer, 2014.

[228] G.D.W. Stetten and R. Klatzky. *Fingertip visual haptic sensor controller*. Google Patents, September 2008. US Patent App. 12/044,138.

[229] Chao-Huai Su, Liwei Chan, Chien-Ting Weng, Rong-Hao Liang, Kai-Yin Cheng, and Bing-Yu Chen. NailDisplay: Bringing an always available visual display to fingertips. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1461–1464, New York, NY, USA, 2013. ACM.

[230] J Sudol, O Dialemah, C Blanchard, and T Dorcey. Looktel: A comprehensive platform for computer-aided visual assistance. In *Proceedings of workshop on Computer Vision Applications for the Visually Impaired*, 2010.

[231] J. Sun. *Ring pointing device*. Google Patents, December 2006. US Patent 7,145,549.

[232] Symbol. WSS 1040: Wearable scanning system, 2000.

[233] W H Talbot, I Darian-Smith, H H Kornhuber, and V B Mountcastle. The sense of flutter-vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. *Journal of Neurophysiology*, 31(2):301–334, 1968.

[234] Emi Tamaki and Ken Iwasaki. A half-implant device on fingernails. In *Proceedings of the 5th Augmented Human International Conference*, AH '14, pages 10:1–10:4, New York, NY, USA, 2014. ACM.

[235] Tao Ma. Phone ring, December 2007.

[236] Thanko. Flick finger mouse, 2013.

[237] Toshiaki Matsushima, Katsuhiro Kanamori, and Sadamu Ohteru. Automated recognition system of musical score (vision system of the WABOT-2). , 3(4):354–361, May 1985.

[238] K. Tsukada and Yasumura. Ubi-finger: Gesture input device for mobile use. In *Proceedings of the APCHI'02*, volume 1, pages 388–400, 2002.

[239] KOJI Tsukada and MICHIAKI Yasumura. Ubi-finger: a simple gesture input device for mobile and ubiquitous environment. *Journal of Asian Information, Science and Life (AISL)*, 2(2):111–120, 2004.

[240] Katia Vega and Hugo Fuks. Beauty technology: Muscle based computing interaction. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '13, pages 469–474, New York, NY, USA, 2013. ACM.

[241] Ramiro Velázquez. Wearable assistive devices for the blind. In Aimé Lay-Ekuakille and SubhasChandra Mukhopadhyay, editors, *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment*, volume 75 of *Lecture Notes in Electrical Engineering*, pages 331–349. Springer Berlin Heidelberg, 2010.

[242] Victor Soto. iRing controls your iPod, September 2007.

[243] Viktor Linder. iSeeNotes - sheet music OCR!, November 2014.

[244] T. Watanabe and H. Iwasaki. Wearable finger dual band antenna for BAN. In *Antennas and Propagation in Wireless Communications (APWC), 2012 IEEE-APS Topical Conference on*, pages 51–54, September 2012.

[245] L.S. Weinblatt. *Finger-mounted computer interface device.* Google Patents, August 1999. US Patent 5,940,066.

[246] Mark Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–104, 1991.

[247] Mark Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84, 1993.

[248] Pierre Wellner. Interacting with paper on the digitaldesk. *Communications of the ACM*, 36(7):87–96, 1993.

[249] Julia Werner, Reto Wettach, and Eva Hornecker. United-pulse: Feeling your partner's pulse. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '08, pages 535–538, New York, NY, USA, 2008. ACM.

[250] F. Winberg and J. Bowers. Assembling the senses: towards the design of cooperative interfaces for visually impaired users. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 332–341. ACM, 2004.

[251] Katrin Wolf. Ubiquitous grasp interfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '13, pages 377–378, New York, NY, USA, 2013. ACM.

[252] Katrin Wolf, Robert Schleicher, Sven Kratz, and Michael Rohs. Tickle: A surface-independent interaction technique for grasp interfaces. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '13, pages 185–192, New York, NY, USA, 2013. ACM.

[253] Y. Xiong. *Finger worn and operated input device and method of use*. Google Patents, November 2003. US Patent App. 10/146,572.

[254] Yoji Yamada, Tetsuya Morizono, Shuji Sato, Takahiro Shimohira, Yoji Umetani, Tetsuji Yoshida, and Shigeru Aoki. Proposal of a SkilMate finger for EVA gloves. In *International Conference on Robotics and Automation*, volume 2, pages 1406–1412, 2001.

[255] Yusuke Yamamoto, Hideaki Uchiyama, and Yasuaki Kakehi. onNote: Playing Printed Music Scores As a Musical Instrument. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 413–422, New York, NY, USA, 2011. ACM.

[256] X.D. Yang, T. Grossman, D. Wigdor, and G. Fitzmaurice. Magic finger: always-available input through finger instrumentation. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 147–156. ACM, 2012.

[257] Xing-Dong Yang, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. Magic Finger: Always-available Input Through Finger Instrumentation. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 147–156, New York, NY, USA, 2012. ACM.

[258] Yang Hai and Mingshuo Zhang. Ring of security, July 2007.

[259] Debra Yarrington and Kathleen McCoy. Creating an Automatic Question Answering Text Skimming System for Non-visual Readers. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, pages 279–280, New York, NY, USA, 2008. ACM.

[260] Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. Current and Future Mobile and Wearable Device Use by People with Visual Impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3123–3132, New York, NY, USA, 2014. ACM.

[261] Chucai Yi and Yingli Tian. Assistive text reading from complex background for blind persons. In *Camera-Based Document Analysis and Recognition*, pages 15–28. Springer, 2012.

[262] S.M. Youssef and R.M. Salem. Automated barcode recognition for smart identification and inspection automation. *Expert Systems with Applications*, 33(4):968–977, 2007.

[263] K. Zawrotny, A. Craig, D. Weiser, R. Klatzky, and G. Stetten. Fingertip vibratory transducer for detecting optical edges using regenerative feedback. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2006 14th Symposium on*, pages 373–374, March 2006.

[264] Boning Zhang, Yiqiang Chen, Yueliang Qian, and Xiangdong Wang. A ring-shaped interactive device for large remote display and mobile device control. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 473–474, New York, NY, USA, 2011. ACM.

[265] C. Zhang, J. Wang, S. Han, M. Yi, and Z. Zhang. Automatic real-time barcode localization in complex scenes. In *Image Processing, 2006 IEEE International Conference on*, pages 497–500. IEEE, 2006.

[266] Thomas G Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A hand gesture interface device. In *ACM SIGCHI Bulletin*, volume 18, pages 189–192. ACM, 1987.

[267] M.M. Zloof. *Hand worn remote computer mouse.* Google Patents, February 1996. US Patent 5,489,922.

[268] Amit Zoran, Roy Shilkrot, Pragun Goyal, Pattie Maes, and Joseph Paradiso. The wise chisel: The rise of the smart handheld tool. *Pervasive Computing, IEEE*, 2014.

[269] Amit Zoran, Roy Shilkrot, Suranga Nanayakkara, and Joseph Paradiso. The hybrid artisans: A case study in smart tools. *To appear in ACM Transactions on Computer-Human Interaction (ToCHI)*, 2014.

[270] Amit Zoran, Roy Shilkrot, and Joseph Paradiso. Human-computer interaction for hybrid carving. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 433–440. ACM, 2013.