AWARD NUMBER: W81XWH-14-1-0617

TITLE: HandSight: Supporting Everyday Activities Through Touch-Vision

PRINCIPAL INVESTIGATOR: Rama Chellappa

CONTRACTING ORGANIZATION: University of Maryland, College Park, MD

REPORT DATE: December 2019

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2019	Final	30Sep2014-29Sep2019

4. TITLE AND SUBTITLE

HandSight: Supporting Everyday Activities Through Touch-Vision

5a. CONTRACT NUMBER
W81XWH-14-1-0617

5b. GRANT NUMBER
MR130161

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jon E. Froehlich, Leah Findlater, Rama Chellappa

E-Mail: chella@umd.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Maryland
3112 Lee Building MD-005
College Park MD 20742-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We built and evaluated a new system called *HandSight* to support activities of daily living (ADLs) for blind and low vision people by sensing and feeding back non-tactile information about the physical world *as it is touched*. HandSight consists of tiny cameras and micro-haptic actuators integrated into one or more fingers, computer vision algorithms to support inference and recognition, and a smartwatch for processing, power, and speech output. The two high-level goals were: first, to develop the basic building blocks of an extensible platform that could support a range of ADL applications. Second, to explore and demonstrate the potential of HandSight through three proof-of-concept applications: reading, dressing, and technology access. Research contributions include functional prototypes, computer vision algorithms to extract attributes of the physical world (e.g., text, color, texture) and to support on-body interaction for mobile technology accessibility, haptic feedback techniques to guide the user's finger/hand, empirical data from user studies with blind and low vision users of all of the aforementioned contributions, and design recommendations for mobile and wearable computer-vision-augmented touch systems.

15. SUBJECT TERMS

Blind, low vision, wearable computing, computer vision, vision-augmented touch, activities of daily living (ADLs), assistive technology, haptics,

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Unclassified	161	USAMRMC
Unclassified	Unclassified	Unclassified			19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

	Page
1. Introduction	3
2. Keywords	3
3. Accomplishments	3
4. Impact	7
5. Changes/Problems	8
6. Products	8
7. Participants & Other Collaborating Organizations	10
8. Special Reporting Requirements	10
9. Appendices	10

1. INTRODUCTION

Over one million US Veterans have visual impairments (VI) that affect their ability to perform activities of daily living—a number that continues to rise as the US Veteran population ages. While previous research has explored combining mobile cameras and computer vision to support people with VI for at-a-distance information tasks such as navigation, facial recognition, and spatial perception, they fail to support proximal information accessed through touch. In our research, we are pursuing a fundamentally new approach: a vision-augmented touch system called *HandSight* aimed at supporting activities of daily living (ADLs) by sensing and feeding back non-tactile information about the physical world *as it is touched*. HandSight consists of tiny CMOS cameras ($1\times1mm^2$) and micro-haptic actuators integrated into one or more fingers, computer vision and machine learning algorithms to support fingertip-based sensing, and a smartwatch for processing, power, and speech output. Since touch is a highly attuned means of acquiring information for people with VI, we hypothesize that collocating the camera with the touch itself will enable new and intuitive assistive applications. Imagine, for example, touching a printed page and receiving speech output while haptic cues guide your fingers along the text, touching a piece of clothing to hear suggestions for coordination, or even picking up a bell pepper and instantly recognizing that it is red or green. In the next sections, we articulate the key objectives and accomplishments of our work.

2. KEYWORDS

Blind, low vision, wearable computing, computer vision, vision-augmented touch, activities of daily living (ADLs), assistive technology, haptics, real-time OCR

3. ACCOMPLISHMENTS

What were the major goals of the project?

This project had two high-level goals:

First, to develop the basic building blocks of an extensible HandSight platform that would support a range of activities of daily living (ADL) applications. This includes: (i) *developing computer vision and machine learning algorithms* to support basic sensing needs from a finger-mounted camera (*e.g.*, optical character recognition, surface and color classification, gesture recognition); and (ii) investigating physical design and feedback modalities (*e.g.,* augmenting one *vs.* both index fingers, using directional haptic feedback to guide the user's touch).

Second, to explore and demonstrate the potential of HandSight through three proof-of-concept applications: reading, dressing, and technology access. Our goal was to demonstrate the benefits of HandSight both for blind and low vision users. The VA estimates that 157,000 US Veterans are legally blind, and more than *one million* have low vision. Unobtrusively reading text, for example, holds significant promise for both populations.

The Statement of Work includes the following tasks:

- **Task 1:** Develop responsive haptic feedback
- **Task 2:** Develop physical form factor
- **Task 3:** Develop sensing algorithms and evaluate using controlled datasets
- **Task 4:** Design, build, and evaluate a reading application
- **Task 5:** Design, build, and evaluate a dressing application
- **Task 6:** Design, build, and evaluate technology access application

What was accomplished under those goals?

The accomplishments of this work are embodied in 15 journal, conference and workshop publications and 2 PhD dissertations (Uran Oh and Lee Stearns). Here, we summarize accomplishments specifically related to each of the six tasks from the SOW before summarizing additional overarching accomplishments.

Task 1: Develop responsive haptic feedback

Toward developing responsive haptic feedback, we first conducted a controlled lab experiment to evaluate the impact of directional wrist-based vibro-motor feedback on hand movement, comparing lower-fidelity (4-motor) and higher-fidelity (8-motor) wristbands. This work did not focus on blind and low vision users and our findings do not necessarily generalize to those populations (see follow-up study, next). Instead, in this initial study, twenty blindfolded participants completed a series of trials, which consisted of interpreting a haptic stimulus and executing a 2D directional movement on a touchscreen. We compared the two conditions in terms of movement error and trial speed, but also analyze the impact of specific directions on performance. Our results showed that doubling the number of haptic motors reduces directional movement error but not to the extent expected. We also empirically derived an apparent lower bound in accuracy of ~25° in interpreting and executing on the directional haptic signal. This work was published as Hong et al., GI 2016.

In follow-up work, we evaluated wrist-worn haptics as a directional hand guide with blind and low vision participants. We implemented and evaluated the following haptic wristband variations: (1) four versus eight vibromotor designs; (2) vibration from only a single motor at a time versus from two adjacent motors using interpolation. To evaluate our designs, we conducted two studies: Study 1 (N=13, 2 blind) showed that participants could non-visually find targets and trace paths more quickly and accurately with single-motor feedback than with interpolated feedback, particularly when only four motors were used. Study 2 (N=14 blind or low vision participants) found that single-motor feedback with four motors was faster, more accurate, and most preferred compared to similar feedback with eight motors. We derived implications for the design of wrist-worn directional haptic feedback and discuss future work. This work was published as Hong et al., ASSETS 2017.

Task 2: Develop physical form factor

The accomplishments related to this task are subsumed in the outcomes listed under Tasks 4, 5, and 6.

Task 3: Develop sensing algorithms and evaluate using controlled datasets

The accomplishments related to this task are subsumed in the outcomes listed under Tasks 4, 5, and 6.

Task 4: Design, build, and evaluate a reading application

Our work on designing, building, and evaluating a reading application consists of an initial proof-of-concept prototype (Stearns et al., ACVR 2014), studies on reading performance with our system (Stearns et al., ACVR 2014; Stearns et al., TACCESS 2016), and extension of the system and user studies to support low vision users (Stearns et al., ASSETS 2017 and ASSETS 2018).

We first built and conducted a pilot evaluation of a finger-worn proof-of-concept system for reading printed text (ACVR 2014). In a user study with three blind and one low vision user, we explored how to continuously guide a user's finger across text using three feedback conditions (haptic, audio, and both). Though preliminary, our results showed that participants valued the ability to access printed material, and that, in contrast to previous findings, audio finger guidance may result in the best reading performance.

In a more comprehensive follow-up effort, we investigated the effectiveness of finger-based sensing and feedback for reading printed text, by conducting a controlled laboratory experiment with 19 blind participants, comparing audio and haptic directional finger guidance within an iPad-based testbed. As a small follow-up, we asked four of those participants to return and provide feedback on a preliminary wearable prototype. Findings from the controlled experiment showed similar performance between haptic and audio directional guidance, although audio may offer an accuracy advantage for tracing lines of text. Subjective feedback also highlighted trade-offs between the two types of guidance, such as the interference of audio guidance with speech output and the potential for desensitization to haptic guidance. While several participants appreciated the direct access to layout information provided by finger-based exploration, important concerns also arose about ease of use and the amount of concentration required. The paper where these results were published (TACCESS 2016) also includes a discussion on the effectiveness of finger-based reading for blind users and potential design improvements to the HandSight prototype.

We then extended this reading work to low vision (Stearns et al., ASSETS 2017 and 2018). Because low vision users often rely on visual information, here we incorporated an augmented reality (AR) approach that could enhance visual

capabilities could make text and other fine details to make them more accessible for low vision users. We followed an iterative design process with feedback and suggestions from seven low vision participants, designing and testing different camera form factors (finger-worn and smartphone-based) and AR magnification ideas using the Microsoft HoloLens. Participants identified several advantages to the concept of head-worn magnification (e.g., portability, privacy, ready availability), and to our AR designs in particular (e.g., a more natural reading experience and the ability to multitask). We discussed the strengths and weaknesses of this AR magnification approach and summarize lessons learned throughout the process.

Task 5: Design, build, and evaluate a dressing application

To adapt the system developed in Task 4 for a dressing application, we focused primarily on algorithmic contributions. First (Medeiros et al., ASSETS 2017), we investigated clothing color and visual texture recognition using images from a finger-mounted camera to support people with visual impairments. Our approach mitigates issues with distance and lighting that can impact the accuracy of existing color and texture recognizers and allows for easy touch-based interrogation to better understand clothing appearance. We classify image textures by combining two off-the-shelf techniques commonly used for object recognition achieving 99.4% accuracy on a dataset of 520 clothing images across 9 texture categories. Then (Stearns et al., ASSETS 2018), to improve scalability and robustness, we collected a dataset of fabric images from online sources and apply transfer learning to train an end-to-end deep neural network to recognize visual patterns. This new approach achieves 92% accuracy in a general case and 97% when tuned for images from a finger-mounted camera.

Task 6: Design, build, and evaluate technology access application

Our technology access work focused primarily on on-body interactions but also explored multi-modal information access more generally.

The technical underpinnings of this work are inspired by biometrics research. First, in our ICPR 2016 paper, we explored the application of skin-based biometrics to our problem. Skin-based biometrics rely on the distinctiveness of skin patterns across individuals for identification. We first investigated whether small image patches of the skin can be localized on a user's body, determining not “who?” instead “where?” Applying techniques from biometrics and computer vision, we introduced a hierarchical classifier that estimates a location from the image texture and refines the estimate with keypoint matching and geometric verification. To evaluate our approach, we collected 10,198 close-up images of 17 hand and wrist locations across 30 participants. Within-person algorithmic experiments demonstrated that an individual's own skin features can be used to localize their skin surface image patches with an F_1 score of 96.5%. As secondary analyses, we assessed the effects of training set size and between-person classification. Our paper also discussed the strengths and limitations of our approach and evaluation methods as well as implications for future applications using a wearable camera to support touch-based, location-specific taps and gestures on the surface of the skin.

We then focused on building an interactive system for mobile access using on-body interactions. While past work had explored a variety of approaches such as wearable depth cameras, bio-acoustics, and infrared reflectance (IR) sensors, these systems did not instrument the gesturing finger, did not easily support multiple body locations, and had not been evaluated with blind and low vision users (our goal). In this IMWUT 2017 paper, we introduced *TouchCam*, a finger wearable to support location-specific, on-body interaction. TouchCam combines data from infrared sensors, inertial measurement units, and a small camera to classify body locations and gestures using supervised learning. We empirically evaluated TouchCam's performance through a series of offline experiments followed by a realtime interactive user study with 12 blind and visually impaired participants. In our offline experiments, we achieved high accuracy (>96%) at recognizing coarse-grained touch locations (e.g., palm, fingers) and location-specific gestures (e.g., tap on wrist, left swipe on thigh). The follow-up user study validated our real-time system and helped reveal tradeoffs between various on-body interface designs (e.g., accuracy, convenience, social acceptability). Our findings also highlighted challenges to robust input sensing for visually impaired users and suggest directions for the design of future on-body interaction systems.

Turning from algorithmic advances and systems building, we then focused on user studies and specifically the use of our system to support microinteractions. For blind screenreader users, today's mobile devices, while reasonably accessible, are not necessarily efficient. This inefficiency may be especially problematic for *microinteractions*, which are brief but high-frequency interactions that take only a few seconds for sighted users to complete (e.g., checking the weather or for new messages). One potential solution to support efficient non-visual microinteractions is on-body input, which appropriates the user's own body as the interaction medium. We thus addressed two related research questions: How well are microinteractions currently supported for blind and low vision users? How should on-body interaction be designed to best support microinteractions for these users? We conducted two studies: (1) an online survey to compare current

microinteraction use between blind and low vision users and sighted users (N=117); and (2) an in-person study where 12 screenreader users qualitatively evaluated a real-time on-body interaction system that provided three contrasting input designs. Our findings suggested that efficient microinteractions are not currently well-supported for blind and low vision users, at least using manual input, which highlights the need for new interaction approaches. On-body input offers this potential and the qualitative evaluation revealed tradeoffs with different on-body interaction techniques in terms of perceived efficiency, learnability, social acceptability, and ability to use on the go.

Finally, we examined a more general question of visual information access and computer vision tools, we addressed the lack of tool support for blind and low vision users in creating visual content. In this preliminary exploration, published at ASSETS 2019 (Potluri et al.), we discussed the design and preliminary exploration of a multi-modal and accessible approach for BVI developers to edit visual layouts of webpages while maintaining visual aesthetics.

Overarching accomplishments

The following accomplishments span multiple if not all of the tasks above. One of our initial accomplishments was to publish an overview of the project at ASSETS 2015 (Findlater et al.) to generate interest in the work and spur discussion with relevant research and end user communities.

In the last year of the project, we also published two reflective papers. The first of these papers focused on head-mounted displays and accessibility—head-mounted displays being a component of some of our HandSight project work to support low vision users. We outlined (Goodman et al., CHI 2019 workshop) key social acceptability challenges that will need to be addressed for HMDs to achieve widespread success as an assistive technology. Specifically, we detailed: (1) tensions between the evolving societal perceptions of HMDs and concerns of stigma and misperception that can arise with assistive technologies, (2) the need to provide assistive support without undue attentional cost and while maintaining expectations for person-to-person communication, and (3) issues of balancing the sensing needs of many of these HMD- based AT systems with the privacy needs of non-wearers. Through this synthesis, we identified open questions and provided guidance for future researchers in this space.

The second paper more broadly examined AI fairness systems that arise with the class of systems that HandSight falls into: AI-based systems that augment sensory abilities (presented at an ASSETS 2019 workshop and in a SIGACCESS 2020 article). Unlike many other AI-enabled technologies, these systems provide information that is already available to non-disabled people. In this article, we discussed unique AI fairness challenges that arise in this context, including accessibility issues with data and models, ethical implications in deciding what sensory information to convey to the user, and privacy concerns both for the primary user and for others.

What opportunities for training and professional development has the project provided?

As university professors, the PIs worked closely with undergraduate and graduate students throughout this project. Students led different parts of the project (e.g., hardware design vs. computer vision vs. interaction design) and gained expertise in those areas. The lead PhD students (Lee Stearns, Uran Oh, Jonggi Hong) also helped to supervise undergraduate students. All students gained technical skills in their areas, skills related to designing and running studies (*e.g.,* experimental design, analysis, running study sessions themselves), and general research skills (*e.g.,* analyzing data, writing papers).

How were the results disseminated to communities of interest?

Our primary form of sharing results has been through academic research papers. As mentioned in the "products" section below, this project resulted in 15 journal, conference or workshop papers and 2 PhD dissertations. PIs Froehlich and Findlater have also given talks on HandSight to a variety of forums, including the University of Washington, Stanford, Microsoft Research, Intel, and Apple. Industry attendees have included employees at Google, Microsoft, Amazon, and Snap Research. Both of our journal papers (TACCESS and IMWUT) were also presented orally at related conferences (ASSETS and UbiComp). Finally, all project information is available via https://makeabilitylab.cs.washington.edu/.

What do you plan to do during the next reporting period?

Nothing to report.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

This project has made the following intellectual contributions, as presented in our research papers:

- Findlater et al., ASSETS 2015: This poster paper provides an overview of HandSight's hardware and software, as well as a summary of findings from our initial area of exploration—reading printed text. Because this work is at an early stage, we also cover open questions and challenges that will need to be addressed as we expand to other application areas.
- Findlater et al., SIGACCESS 2020: In this workshop paper, we discuss these questions and outline directions where accessibility and AI/ML researchers can work together to address issues of fairness with AI- enabled assistive technologies for sensory augmentation.
- Goodman et al., CHI 2019 workshop: In this paper, to motivate and encourage future work, we outline key social challenges in using HMDs for accessibility.
- Hong et al., GI 2016: The contributions include: (i) empirical evidence that doubling the number of haptic motors reduces directional movement error, and that movement error is greater to the upper-left than in other directions; (ii) identification of an empirically derived maximum threshold of ~25° accuracy using our approach; (iii) design considerations for incorporating directional haptic guidance into a smartwatch band—in particular, for most applications, a four-motor wristband may be sufficient.
- Hong et al., ASSETS 2017: this paper contributes empirical evidence that wrist-worn haptic feedback with a single motor vibrating at once performs better than interpolated feedback, and that embedding four motors in the wristband performs better than eight motors. We also provide design implications for non-visual wrist-worn haptic guidance. Our results have implications for future smartwatch haptic wristband designs.
- Medeiros et al., ASSETS 2017: Our contributions include: (i) extending our finger mounted camera system [6, 10] to support robust color and visual texture recognition accessed through touch, (ii) a small but novel dataset of close-up clothing images captured using this system, and (iii) results of applying a state-of-the-art texture classification approach to our dataset. Because this work is at an early stage, we also cover open questions and challenges that will need to be addressed as we implement an interactive system for visually impaired users.
- Oh et al., ASSETS 2017: The primary contributions of this paper include: (1) identification of the needs, barriers, and strategies for enabling microinteractions for users with visual impairments in comparison to sighted users, and (2) design implications for supporting microinteractions for VI users via on-body input.
- Stearns et al., ACVR 2014: This paper contributes a proof-of-concept system and pilot evaluation of a finger-worn reading system.
- Stearns et al., ICRP 2016: The contributions of this paper include: (i) a robust algorithmic pipeline for recognizing several different locations on the hand from small patches of skin; (ii) classification results for a dataset consisting of 30 individuals, achieving accuracy above 96% on average for within-person experiments; and (iii) analysis of hand distinctiveness and similarities among users, which may impact accuracy and scalability (e.g., between-person training feasibility).
- Stearns et al., TACCESS 2016: The contributions of this paper are: (1) empirical results comparing audio and haptic directional finger guidance for a reading task in terms of user performance and subjective response; (2) the implementation and preliminary evaluation of a real- time proof-of-concept system that combines a small finger-mounted camera and feedback mechanism with efficient computer vision algorithms to read printed text; and (3) design reflections for finger-based reading devices for people who are blind.
- Stearns et al., ASSETS 2017 poster: Our contributions include: (i) a preliminary investigation of a finger-mounted camera with an AR display for viewing printed or other non-tactile information and (ii) a proof-of-concept AR system with three distinct interface options for displaying magnified visual content.
- Stearns et al., IMWUT 2017: The primary contributions of this paper include: (i) two iterations of TouchCam, a novel finger- worn camera system that uses machine learning to detect and recognize on-body location-specific gestures; (ii) a quantitative evaluation of our system's accuracy and robustness across a variety of gestures and body locations; (iii) qualitative observations about the usability and utility of our on-body input approach for users with visual impairments; and (iv) design reflections for on-body gestural interfaces in terms of what locations and gestures can be most reliably recognized across users.
- Stearns et al, ASSETS 2018a: Our contributions include: (i) an exploration of the design space for augmented reality magnification; (ii) proof-of- concept implementations evaluated and refined through iterative design with

low vision users; and (iii) common themes and recommendations that should inform the design of future AR vision enhancement aids for low vision users.
- Stearns et al., ASSETS 2018b: Our contributions include: (i) a novel dataset of 77,052 fabric images across six visual patterns; (ii) results of applying transfer learning to train a convolutional neural network for classifying fabric patterns; (iii) a preliminary system to describe clothing appearance in realtime using images from a finger-mounted camera.
- Potluri et al., ASSETS 2019: The poster paper discusses the design and preliminary exploration of a multi-modal and accessible approach for BVI developers to edit visual layouts of webpages while maintaining visual aesthetics.

What was the impact on other disciplines?
Nothing to report.

What was the impact on technology transfer?
We have received multiple inquiries from potential end users about the availability of our work as a commercial product. However, maturing the research prototypes sufficiently to support commercialization was beyond the scope of this grant.

What was the impact on society beyond science and technology?
Our most recent papers (2019 and 2020) papers discussing tensions with AI, equity and social acceptability of AI-based assistive technologies are contributing to ongoing societal discussions about AI fairness.

5. Changes/Problems

In Fall 2017, PIs Froehlich and Findlater left the University of Maryland to join the University of Washington. Due to the timeline required to transfer the grant from Maryland to Washington, the overall length of the project was extended to September 2019.

6. Products

Journal publications

Stearns, L., Du, R., Oh, U., Jou, C., Findlater, L., Ross, D., Froehlich, J. E. (2016). Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. ACM Transactions on Accessible Computing (TACCESS) 2016. https://doi.org/10.1145/2914793

Stearns, L., Oh, U., Findlater, L., Froehlich, J. E. (2017). TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. ACM IMWUT December 2017. https://doi.org/10.1145/3161416

Other publications, conference papers, and presentations

Findlater, L., Stearns, L., Du, R., Oh, U., Ross, D., Chellappa, R., Froehlich, J. E. (2015). Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. Poster Proceedings of ASSETS 2015. https://doi.org/10.1145/2700648.2811381

Findlater, L., Goodman, S., Zhao, Y., Azenkot, S., and Hanley, M. (2020). Fairness Issues in AI systems that Augment Sensory Abilities. ACM SIGACCESS Accessibility and Computing, (125), 1-1. https://doi.org/10.1145/3386296.3386304

Goodman, S., Jain, D., Froehlich, J.E., Craft, B., Findlater, L. (2019). Social Tensions in Head-mounted Displays for Accessibility. Proceedings of the 1st Workshop on Challenges Using Head-Mounted Displays in Shared and Social Spaces at ACM CHI 2020.

Hong, J., Stearns, L., Cheng, T., Froehlich, J. E., Ross, D., Findlater, L. (2016). Evaluating Angular Accuracy of Wrist-based Haptic Directional Guidance for Hand Movement. Proceedings of GI 2016. https://doi.org/10.20380/GI2016.25

Hong, J., Pradhan, A., Froehlich, J. E., Findlater, L. (2017). Evaluating Wrist-Based Haptic Feedback for Non-Visual Target Finding and Path Tracing on a 2D Surface. Proceedings of ASSETS 2017. https://doi.org/10.1145/3132525.3132538

Medeiros, A., Stearns, L., Findlater, L., Chen, C., Froehlich, J. E. (2017). Recognizing Clothing Colors and Visual Textures Using a Finger-Mounted Camera: An Initial Investigation. Proceedings of ASSETS 2017. https://doi.org/10.1145/3132525.3134805

Oh, U., Stearns, L., Pradhan, A., Froehlich, J. E., Findlater, L. (2017). Investigating Microinteractions for People With Visual Impairments and the Potential Role of on-Body Interaction. Proceedings of ASSETS 2017. https://doi.org/10.1145/3132525.3132536

Potluri, V., He, L., Chen, C., Froehlich, J.E., Mankoff, J. (2019). A Multi-Modal Approach for Blind and Visually Impaired Developers to Edit Webpage Designs. Proceedings of ASSETS 2019. https://doi.org/10.1145/3308561.3354626

Stearns, L., Du, R., Oh, U., Wang, Y., Chellappa, R., Findlater, L., Froehlich, J. E. (2014). The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind. Proceedings of ACVR 2014.

Stearns, L., Oh, U., Cheng, B., Findlater, L., Ross, D., Chellappa, R., Froehlich, J. E. (2016). Localization of Skin Features on the Hand and Wrist From Small Image Patches. Proceedings of ICPR 2016. DOI: 10.1109/ICPR.2016.7899767

Stearns, L., De Souza, V., Yin, J., Findlater, L., Froehlich, J. E. (2017). Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera. Proceedings of ASSETS 2017. http://doi.acm.org/10.1145/3132525.3134812

Stearns, L., Findlater, L., Froehlich, J. E. (2018). Design of an Augmented Reality Magnification Aid for Low Vision Users. Proceedings of ASSETS 2018. https://doi.org/10.1145/3234695.3236361

Stearns, L., Findlater, L., Froehlich, J. E. (2018). Applying Transfer Learning to Recognize Clothing Patterns Using a Finger-Mounted Camera. Proceedings of ASSETS 2018. https://doi.org/10.1145/3234695.3241015

Website(s) or other Internet site(s)

https://jonfroehlich.github.io (PI website)

https://makeabilitylab.cs.washington.edu/project/handsight/ (project website)

http://www.instructables.com/id/HandSight-A-Glove-for-the-Blind-to-Feel-Shapes-an/ (Instructable)

https://youtu.be/VREiWI_38BQ (IMWUT'17 TouchCam video)

https://youtu.be/AWCEDYdwBAc (ACVR'14 HandSight video)

https://youtu.be/i0IDbHGir-8 (ASSETS'18 Visual Magnification video)

Technologies or techniques

Nothing to report.

Inventions, patent applications, and/or licenses

Nothing to report.

Other products

The software and hardware files for TouchCam are open sourced and available on github:

https://github.com/lstearns86/touchcam

7. Participants & Other Collaborating Organizations

This project was a collaboration between the University of Maryland, College Park, the Atlanta VA, and, through a subaward, the University of Washington. Over the period of this grant, the following personnel were paid for their research effort:

David Ross
Jon Froehlich
Leah Findlater
Lee Stearns
Uran Oh
Jonggi Hong
Liang He
Dhruv Jain
Venkatesh Potluri
Manaswi Saha

8. Special Reporting Requirements

Nothing to report.

9. Appendices

Findlater, L., Stearns, L., Du, R., Oh, U., Ross, D., Chellappa, R., Froehlich, J. E. (2015). Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. Poster Proceedings of ASSETS 2015. https://doi.org/10.1145/2700648.2811381

Findlater, L., Goodman, S., Zhao, Y., Azenkot, S., and Hanley, M. (2020). Fairness Issues in AI systems that Augment Sensory Abilities. ACM SIGACCESS Accessibility and Computing, (125), 1-1. https://doi.org/10.1145/3386296.3386304

Goodman, S., Jain, D., Froehlich, J.E., Craft, B., Findlater, L. (2019). Social Tensions in Head-mounted Displays for Accessibility. Proceedings of the 1st Workshop on Challenges Using Head-Mounted Displays in Shared and Social Spaces at ACM CHI 2020.

Hong, J., Stearns, L., Cheng, T., Froehlich, J. E., Ross, D., Findlater, L. (2016). Evaluating Angular Accuracy of Wrist-based Haptic Directional Guidance for Hand Movement. Proceedings of GI 2016. https://doi.org/10.20380/GI2016.25

Hong, J., Pradhan, A., Froehlich, J. E., Findlater, L. (2017). Evaluating Wrist-Based Haptic Feedback for Non-Visual Target Finding and Path Tracing on a 2D Surface. Proceedings of ASSETS 2017. https://doi.org/10.1145/3132525.3132538

Medeiros, A., Stearns, L., Findlater, L., Chen, C., Froehlich, J. E. (2017). Recognizing Clothing Colors and Visual Textures Using a Finger-Mounted Camera: An Initial Investigation. Proceedings of ASSETS 2017. https://doi.org/10.1145/3132525.3134805

Oh, U., Stearns, L., Pradhan, A., Froehlich, J. E., Findlater, L. (2017). Investigating Microinteractions for People With Visual Impairments and the Potential Role of on-Body Interaction. Proceedings of ASSETS 2017. https://doi.org/10.1145/3132525.3132536

Potluri, V., He, L., Chen, C., Froehlich, J.E., Mankoff, J. (2019). A Multi-Modal Approach for Blind and Visually Impaired Developers to Edit Webpage Designs. Proceedings of ASSETS 2019. https://doi.org/10.1145/3308561.3354626

Stearns, L., Du, R., Oh, U., Wang, Y., Chellappa, R., Findlater, L., Froehlich, J. E. (2014). The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind. Proceedings of ACVR 2014.

Stearns, L., Oh, U., Cheng, B., Findlater, L., Ross, D., Chellappa, R., Froehlich, J. E. (2016). Localization of Skin Features on the Hand and Wrist From Small Image Patches. Proceedings of ICPR 2016. DOI: 10.1109/ICPR.2016.7899767

Stearns, L., Du, R., Oh, U., Jou, C., Findlater, L., Ross, D., Froehlich, J. E. (2016). Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. ACM Transactions on Accessible Computing (TACCESS) 2016. https://doi.org/10.1145/2914793

Stearns, L., De Souza, V., Yin, J., Findlater, L., Froehlich, J. E. (2017). Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera. Proceedings of ASSETS 2017. http://doi.acm.org/10.1145/3132525.3134812

Stearns, L., Oh, U., Findlater, L., Froehlich, J. E. (2017). TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. ACM IMWUT December 2017. https://doi.org/10.1145/3161416

Stearns, L., Findlater, L., Froehlich, J. E. (2018). Design of an Augmented Reality Magnification Aid for Low Vision Users. Proceedings of ASSETS 2018. https://doi.org/10.1145/3234695.3236361

Stearns, L., Findlater, L., Froehlich, J. E. (2018). Applying Transfer Learning to Recognize Clothing Patterns Using a Finger-Mounted Camera. Proceedings of ASSETS 2018. https://doi.org/10.1145/3234695.3241015

Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch

Leah Findlater[1], Lee Stearns[2], Ruofei Du[2], Uran Oh[2], David Ross[3], Rama Chellappa[4], Jon E. Froehlich[2]

[1]College of Information Studies, [2]Dept. of Computer Science, [4]Dept. of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742

[3]Atlanta VA R&D Center of Excellence in Vision and Neurocognitive Rehabilitation (CVNR)
1670 Clairmont Road, Decatur, GA 30033

ABSTRACT

The *HandSight* project investigates how wearable micro-cameras can be used to augment a blind or visually impaired user's sense of touch with computer vision. Our goal is to support an array of activities of daily living by sensing and feeding back non-tactile information (*e.g.*, color, printed text, patterns) about an object as it is touched. In this poster paper, we provide an overview of the project, our current proof-of-concept prototype, and a summary of findings from finger-based text reading studies. As this is an early-stage project, we also enumerate current open questions.

Categories and Subject Descriptors

K.4.2 [**Computer and Society**]: Social Issues – Assistive technologies for persons with disabilities

Keywords

Blind; visually impaired; wearable computing; computer vision; vision-augmented touch

1. INTRODUCTION

The majority of work on computer vision to support blind and visually impaired users has focused on at-a-distance information tasks such as navigation (*e.g.*, [1]) and spatial perception (*e.g.*, [3]). Instead, the *HandSight* project explores how computer vision can support proximal information accessed through touch. By sensing and feeding back non-tactile information about the physical world *as it is touched*, our goal is to support an array of activities of daily living (ADLs). This solution could, for example, allow the user to touch and read a printed page via real-time OCR and speech synthesis, touch a piece of clothing to hear suggestions for coordinating items, or pick up a bell pepper and recognize that it is red or green. Since touch is a highly attuned means of acquiring information for people with visual impairments (*e.g.*, [2]), we hypothesize that collocating sensing with the user's touch will enable new and intuitive assistive applications.

HandSight is part of a small but growing body of work on augmenting touch with wearable cameras and computer vision algorithms. Perhaps the most well-known system is *Orcam*, a recently introduced commercial head-mounted camera system designed to recognize objects and read printed text. While live demonstrations have been impressive[1], there is no academic work examining Orcam's effectiveness. This head-mounted camera approach could potentially complement HandSight's finger-mounted sensing. More closely related to HandSight are *Magic Finger* [8] and *FingerReader* [4,5]. Magic Finger uses a finger-mounted camera with the goal of "enabling any surface to act as a touch screen," although the focus is not for people with visual impairments—most problematically, the device covers the fingertip, thus interfering with tactile sensation [8]. FingerReader, in contrast, is a custom finger-mounted device with vibration motors designed to allow visually impaired wearers to read printed text by direct line-by-line scanning with the finger [4,5]. While promising, user evaluation of FingerReader to date has been limited to small qualitative studies (3-4 participants).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
ASSETS '15, October 26-28, 2015, Lisbon, Portugal
ACM 978-1-4503-3400-6/15/10.
http://dx.doi.org/10.1145/2700648.2811381



Figure 1: HandSight uses a 1×1mm² AWAIBA NanEye 2C camera developed for minimally invasive surgeries (endoscopies) that can capture 250×250px images at 44fps. Early form factors also shown.

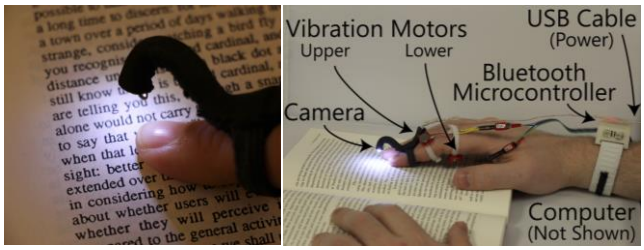
This poster paper provides an overview of HandSight's hardware and software, as well as a summary of findings from our initial area of exploration—reading printed text. Because this work is at an early stage, we also cover open questions and challenges that will need to be addressed as we expand to other application areas.

2. CURRENT PROTOTYPE

Our envisioned system consists of tiny CMOS cameras (1×1mm²) and micro-haptic actuators mounted on one or more fingers, computer vision and machine learning algorithms to support fingertip-based sensing, and a smartwatch for processing, power, and speech output. Figure 1 shows early form factors that we have been exploring, including mounting the camera on a ring or unobtrusively under an acrylic fingernail. Our overarching design goals are that HandSight should: (1) Support touch-based rather than at-a-distance interaction; (2) not hinder normal tactile function, such as covering the fingertip; (3) be easy-to-learn/use (an issue with many sensory aids) through speech and simple haptic feedback; (4) be always-available to allow for seamless transitions to/from real-world tasks; (5) be comfortable & robust to everyday stresses such as water and impact.

The current proof-of-concept prototype consists of a laptop running custom software, external speakers, a finger-mounted camera, and a custom haptic device. The camera is a self-illuminated Awaiba NanEye 2C CMOS camera with an LED ring (~40 fps, 250x250 pixels, 2.4mm diameter; Figure 1a), embedded in an adjustable ring and positioned above the finger. Because our initial application area is reading, this camera position is designed

[1] OrCam TED@NYC. YouTube Video: http://youtu.be/_3XVsCsscyw?t=4m53



(a) Finger-mounted camera (b) Current prototype

Figure 2: Early design for our proof-of-concept prototype with the camera placed for reading printed text. Future designs may be made much smaller, for both the camera mount and the haptic feedback.

to offer a reasonably large field of view for optical character recognition (OCR) on a printed page. Two disc vibration motors (8mm diameter, 3.4mm thick), controlled by an Arduino Pro Micro, provide haptic feedback. The motors are attached to the user's right index finger with separate Velcro rings (Figure 2b), one on top of the finger on the intermediate phalange and one below the finger on the proximal phalange.

3. INITIAL APPLICATION: READING

We have initially explored HandSight to read printed text. Compared to handheld text scanners such as mobile phone applications, a finger-based approach has the potential to mitigate camera framing issues, enable a blind reader to better understand the spatial layout of a document, and provide better control over reading pace. However, such an approach also needs to guide the reader in physically navigating a document and tracing along lines of text. While previous work has proposed audio and haptic directional finger guidance for this purpose, user studies have been limited to 3 or 4 participants [4,5,7]. The most recent of these studies underscores the need for further investigation: all 3 participants found it difficult to read text with FingerReader [5]. This provokes the question: *why*? Are finger-based cameras a viable accessibility solution for reading printed text or simply too challenging to use given issues of accurate text capture and recognition as well as finger guidance?

To further investigate the feasibility of finger-based reading, we conducted two lab studies [6]. The first study included 20 blind participants and compared audio and haptic directional finger guidance during reading tasks. Users traced their finger left to right along each line of text and the system generated text-to-speech output. In this study, we isolated the user experience from the wearable prototype's OCR performance by displaying documents on an iPad (the touchscreen provided accurate sensing of the finger location). Audio guidance played a continuous tone when the user's finger strayed off the current line, with a low pitch indicating the need for downward movement and high pitch indicating upward. For haptic guidance, the lower motor (proximal phalange) vibrated to indicate downward movement and the upper motor indicated the opposite. Participants used both types of guidance in counterbalanced order to read plain text documents and magazine-style documents (with pictures).

The two types of guidance resulted in relatively similar line tracing performance, with speeds ranging from 106–120 WPM across all conditions. Audio, however, may offer an accuracy advantage—with the magazine document participants strayed farther from the line with haptic than with audio (M=15px *vs.* 11 px away from the middle of the line; Z_{19}=-2.374, p=.018, r=.54).

Preference was almost evenly split (11 of 20 participants preferred haptic guidance), reflecting contradictions found in previous research [4,5,7]. Open-ended comments also highlighted the tradeoffs of the two types of guidance. Some participants felt that audio interfered with the speech output: *"You could focus on the audio of the text, and not be listening for other sounds"* (P8), or *"I missed a couple words because I was being distracted by the [audio]"* (P16). Those who preferred the audio reported that haptic guidance required more effort and could lead to desensitization. Finally, while several participants appreciated the direct access to layout information provided with HandSight's exploration mode, important concerns arose overall about ease of use and the amount of concentration required.

In the second study, 4 blind participants used a proof-of-concept wearable prototype (Figure 2b) and KNFB Reader iOS, a state-of-the-art OCR application, to complete reading tasks. The findings highlighted the strengths and weaknesses of the two approaches. HandSight offered additional layout information and more immediate access to text content without the need to worry about framing the complete document within the camera's field of view. But, once the document was successfully scanned, KNFB Reader offered a faster and smoother reading experience, and was preferred by all participants. Further work is needed to assess the impacts of the tradeoffs on document comprehension.

4. DISCUSSION AND OPEN QUESTIONS

The goal of the HandSight project is to support an array of activities of daily living by sensing and feeding back non-tactile information (*e.g.*, color, printed text, visual patterns) about an object as it is touched. We plan to explore several application areas beyond reading; however, many open questions remain. While a finger-based camera is always available and may mitigate camera-framing issues that can occur with mobile devices (*e.g.*, smartphones), wearables introduce practical issues of robustness and social acceptability. Future prototypes will be smaller and less obtrusive, leveraging our tiny camera form factor. We are also studying how to incorporate smartwatches more effectively, such as offloading haptic feedback to the band of the watch. Still, whether blind users are interested in wearing a finger-mounted device for accessibility is an open question.

5. ACKNOWLEDGEMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

6. REFERENCES

[1] Dakopoulos, D. and Bourbakis, N.G. Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. *IEEE SMC, Part C (Applications and Reviews) 40*, 1 (2010), 25–35.

[2] Goldreich, D. and Kanics, I.M. Tactile Acuity is Enhanced in Blindness. *J. Neurosci. 23*, 8 (2003), 3439–3445.

[3] Khambadkar, V., Folmer, E. GIST: A Gestural Interface for Remote Nonvisual Spatial Perception. *Proc. ACM UIST'13*, (2013), 301–310.

[4] Shilkrot, R., Huber, J., Liu, C., Maes, P., and Chandima, N.S. FingerReader: a wearable device to support text reading on the go. *CHI '14 Extended Abstracts*, (2014), 2359–2364.

[5] Shilkrot, R., Huber, J., Wong, M.E., Maes, P., and Nanayakkara, S. FingerReader: a wearable device to explore printed text on the go. *Proc. ACM CHI'15*, (2015), 2363–2372.

[6] Stearns, L., Du, R., Oh, U., et al. Evaluating Haptic and Auditory Directional Guidance to Assist Blind Persons in Reading Printed Text Using Finger-Mounted Cameras. *In submission*.

[7] Stearns, L., Du, R., et al. The design and preliminary evaluation of a finger-mounted camera and feedback system to enable reading of printed text for the blind. *ACVR'14 Workshop at ECCV'14*, (2014).

[8] Yang, X.-D., Grossman, T., Wigdor, D., and Fitzmaurice, G. Magic finger: always-available input through finger instrumentation. Proc. ACM UIST '12, (2012), 147–156.

Fairness Issues in AI Systems that Augment Sensory Abilities

Leah Findlater, Steven Goodman
Human Centered Design & Engineering | DUB Group
University of Washington
Seattle, WA, USA
{leahkf, smgoodmn}@uw.edu

Yuhang Zhao
Information Science,
Cornell Tech
Cornell University
New York, NY, USA
yz769@cornell.edu

Shiri Azenkot, Margot Hanley
Jacobs Technion-Cornell Institute, Cornell Tech
Cornell University
New York, NY, USA
shiri.azenkot@cornell.edu

ABSTRACT
Systems that augment sensory abilities are increasingly employing AI and machine learning (ML) approaches, with applications ranging from object recognition and scene description tools for blind users to sound awareness tools for d/Deaf users. However, unlike many other AI-enabled technologies, these systems provide information that is already available to non-disabled people. In this paper, we discuss unique AI fairness challenges that arise in this context, including accessibility issues with data and models, ethical implications in deciding what sensory information to convey to the user, and privacy concerns both for the primary user and for others.

Author Keywords
Accessibility; AI fairness; sensory abilities.

INTRODUCTION
An increasing number of AI-enabled assistive technologies leverage advances in computer vision, speech recognition, auditory scene analysis, and other applications of machine learning to sense and interpret information about the real world for users with disabilities. Active areas of work range from navigation and object recognition tools for blind and low vision users (e.g., [7, 13, 21]) to sound sensing and feedback systems for deaf and hard of hearing users (e.g., [16, 19]).

As the underlying AI advances, these systems will push the boundaries of (dis)ability for people with sensory disabilities. They will ensure that a Deaf person hears a knock on the door or, more fundamentally, shape the way a blind person perceives and engages with people around her. However, unlike many other AI-enabled technologies, these systems provide information that is *already* available to non-disabled people. This goal of equal access raises unique fairness issues for AI-enabled assistive technology: How can we ensure the data and models themselves are accessible? What are the ethical implications of an AI system that determines what information is conveyed to the user about their environment? How can the privacy issues of always-on sensing be mitigated both for the primary user and for others in a larger social sphere?

In this workshop paper, we discuss these questions and outline directions where accessibility and AI/ML researchers can work together to address issues of fairness with AI-enabled assistive technologies for sensory augmentation.

DATA AND MODEL (IN)ACCESSIBILITY
Since the data used by AI-enabled assistive technologies is inherently not accessible to its primary users, several fairness issues related to data and model accessibility arise. These include questions of how to accessibly explain the system's decision-making process, how to help the user provide effective input to the system, and how to enable the user to verify recognition results.

Observing how AI models interpret input/data is critical for developing appropriate expectations, fostering trust in the system, and making responsible decisions based on its output. With the European Union's "right to explanation" about automated decision-making coming into force in 2018 [10], policy is beginning to reflect a push for transparency in model-based decision-making. Research shows that clarity of an AI model and its limitations can foster trust in the system and drive its continued use [6, 18]. However, in the context of AI-based sensory tools, the data is inherently inaccessible to the target users of the tools, making it difficult for those users to understand the model and thus achieve trust and sound decision-making. For example, one approach to explain an image classification model is to highlight the locations in the image that the model finds to be the most salient or distinguishable (e.g., [18]). These visual explanations, however, are not accessible to blind people, which limits *who* can participate in the development and/or evaluation of the model.

Inaccessible data (and explanations) can also limit the user's ability to personalize their AI system. Kacorri, et al. [12] argue that AI models constrained to an individual user's accessibility needs can work more effectively than more

To appear at:
Findlater, L., Goodman, S., Zhao, Y., Azenkot, S., and Hanley, M. 2019. Fairness issues in AI systems that augment sensory abilities. *ACM ASSETS Workshop on AI Fairness for People with Disabilities*, 4 pages.

Copyright retained by authors.

general models. As a case study [13], that same research team explored how visually-impaired people might train personalized object recognition systems with their own snapshots, indicating a need for feedback on the quality of the collected training data (e.g., lighting conditions, number of photos). A challenge of accessible model training, then, is implementing feedback on the quality of collected data in a modality that is accessible to the user. Our sound awareness work for deaf and hard-of-hearing users (e.g., [8]) offers another example, because supporting a d/Deaf user in personalizing their sound classifier requires non-auditory feedback on the quality of each training example. But without the user knowing the characteristics of the sound they are interested in (e.g., possible waveform shape, if the sound is even occurring), how can this feedback be effectively implemented?

Finally, when people with disabilities use AI systems to provide information about their surroundings, inaccurate recognition results can mislead the user, magnifying their vulnerability and even harming their safety (e.g., recognizing a stranger as a friend). This high cost makes it particularly problematic when people with sensory disabilities are not able to determine the accuracy of a system's recognition result. As a consequence, users may place too much trust in the system (e.g., tending to trust inaccurate image description results [14]), or may resort to workarounds that have other negative consequences such as limiting independence (e.g., asking sighted people to double check the system's output) [22]. Thus, one important research question is how to support people with sensory disabilities in independently determining the quality and uncertainty (e.g., confidence level) of AI sensing results.

DECISION-MAKING IN AI-BASED SENSING

Because AI-based sensing systems can strongly influence a user's understanding of the world around them, the seemingly simple task of describing sensory input through language is highly complex. In describing a visual or auditory scene, how does a human or AI system decide which details to include and which to leave out?

Consider facial recognition technology for blind and low vision users. Myriad decisions arise in how to implement such a system, with cascading social consequences for the user themselves, and the potential to perpetuate social bias and identify members of marginalized groups without their consent. For example, should such an app only identify by name people the user already knows well, more remote acquaintances, online friends they have never met in person, strangers, celebrities? Besides names, what other information should be conveyed? What about sensitive social categories that sighted people are constantly judging based on visual cues, such as age, gender, and race? On the one hand, an equal access argument can be made that blind users should be able to have this information as well, but this argument is in tension with known issues of bias in AI systems (e.g., [4]). There is also the question of who gets to decide what labels should be used—members of the community being identified, big tech companies, end users?—which reveals the underlying fairness-related tensions at play.

For a less sensitive example, consider a sound awareness system: even in a quiet office, sounds can include a ticking clock, chairs squeaking, typing, a quiet phone conversation, cars and birds outside, the hum of the heating system, and so on. Understandably, a recent study of user needs for sound awareness systems showed that most deaf and hard of hearing participants wanted the system to filter out unimportant sounds, such as ambient background sounds [8]. To what extent can a deaf user trust an AI-based sensing system to notify them of important sounds and to filter out unimportant ones? Who should determine what is an important sound and how that sound is described?

Even this seemingly-mundane use case is loaded with consequences of multiple layers of human decisions, amplified through AI, that shape the way people with sensory disabilities perceive and make sense of the world around them. Who decides what labels to use when labeling training data? Who applies those labels and what biases might they have? Who decides what datasets and modeling approaches should be incorporated into the final system? And how should we present the model's output to the user? Today, we have little visibility into who makes these decisions and how they are made. "Datasheets for Datasets" [9] aims to address this gap by creating prototypes of "datasheets"—forms that accompany every dataset, fostering a level of critical thinking during data collection and transparency afterward. While this is a step in the right direction, it only begins to address the first two layers of decision-making. Researchers should study the entire decision-making pipeline, and consider following Gebru et al.'s approach to improve transparency.

INDIVIDUAL AND SOCIETAL PRIVACY ISSUES

Augmenting sensory abilities raises privacy issues both for the primary user of the system and for others, particularly when these AI-based systems employ always-on sensors like wearable cameras or live microphones. As a society we will need to evolve policies and practices that we feel are just yet that also balance the right to privacy.

For the primary user of the system, privacy issues arise from the risk of inadvertently sharing personally sensitive data and from pressure the user may feel to relinquish their privacy in order to achieve the benefits of using the system (e.g., gaining a description of an image or sound clip). A classic example is smartphone-based object recognition for blind users—whether fully AI-driven (e.g., Microsoft SeeingAI [15]) or crowd- and/or human-service- driven (e.g., VizWiz [3], Aira [2]). The recent publication of a VizWiz usage dataset of pictures with private information includes examples ranging from prescription and credit card information to family photos and pregnancy test results [11]. Creators of AI-based sensing systems are responsible for

mitigating these privacy risks and for appropriately conveying risks to the user.

Complex privacy concerns also arise when considering the broader social context in which these systems will be used. Conversation partners' and bystanders' attitudes are evolving about the use of always-on, AI-based sensing in public contexts, impacted by whether those individuals realize that the device is providing assistance to the wearer, changing societal norms around wearable technologies in general, and legal and policy implications. For example, Profita et al. [17] showed that third-party onlookers were more accepting of public head-mounted display use (with an onboard camera) when they thought the wearer of the device had a disability, and even more so when they knew the device was being used for an assistive purpose. Ahmed et al. [1] investigated the use of HMDs in a workplace environment, examining what information co-workers were comfortable sharing via these devices. A main takeaway of Ahmed et al.'s study emphasizes the intersection with accessibility: sighted co-workers were more willing to share personal information (e.g., demographics, contact information) with visually impaired colleagues than with other sighted co-workers.

Finally, there are legal and surveillance implications for AI-based assistive sensing systems. Several US states have all-party consent laws, where everyone involved in a conversation must consent to being audio recorded. Lawmakers are also grappling with how to regulate rapidly emerging technologies such as facial recognition software. In 2019, for example, San Francisco banned the use of facial recognition technology by police and other agencies, a response to concerns about the unprecedented opportunities that these technologies offer for government surveillance [5]. For Deaf users, however, audio recording may enable useful sound awareness features, and for blind users, facial recognition software may be useful for getting an understanding of who is around or who is approaching, in both cases providing information that may be seen as equitable to what a non-disabled individual already has access to. How these complexities play out with assistive sensing systems must be carefully considered. One potential path is for there to be "assistive use" exceptions for some of these technologies that are similar to legal exceptions for service animals, where service animals can accompany people with disabilities in all areas that members of the public are normally allowed to go, even places where animals are otherwise not allowed (e.g., the ADA in the US [20]).

CONCLUSION

We have outlined AI fairness issues that arise with AI-enabled assistive technology for people with sensory disabilities, highlighting open challenges for the accessibility community as well as directions that will be best addressed through close collaboration between AI and accessibility researchers. While the data itself is inherently inaccessible in an assistive sensing context, we can work with AI/ML researchers to address accessibility issues related to model explainability and personalization. Studying the entire decision-making pipeline for AI-based assistive sensing systems is also important, and following Gebru et al.'s [9] approach to improve transparency may be useful here. Finally, as a community, we should establish guidelines to help people reach decisions that balance privacy and fairness considerations for AI sensing, and actively contribute to policy decisions regarding AI, such as the "right to explanation" and privacy protections.

ACKNOWLEDGMENTS

This work was supported in part by the Office of the Assistant Secretary of Defense for Health Affairs under award W81XWH-14-1-0617 and by the National Science Foundation under award IIS-1763199. We also thank Hal Daumé III for early discussion related to this paper.

REFERENCES

1. Ahmed, T., Kapadia, A., Potluri, V., & Swaminathan, M. 2018. Up to a limit?: Privacy concerns of bystanders and their willingness to share additional information with visually impaired users of assistive technologies. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3), 89.
2. Aira. 2018. Retrieved July 2, 2019 from https://aira.io/
3. Bigham, J. P., Jayant, C., Ji, H., et al. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*, 333-342. ACM.
4. Buolamwini, J., & Gebru, T. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the Conference on Fairness, Accountability and Transparency*, 77-91.
5. Conger, K, Fausset, R. and Kovaleski, S.F. 2019, May 14. San Francisco bans facial recognition technology. *New York Times*. Retrieved July 1, 2019 from https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html
6. Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697-718.
7. Fiannaca, A., Apostolopoulous, I., & Folmer, E. 2014. Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces. *Proceedings of the 16th international ACM SIGACCESS Conference on Computers & Accessibility*, 19-26. ACM.
8. Findlater, L., Chinh, B., Jain, D., Froehlich, J., Kushalnagar, R., & Lin, A. C. 2019. Deaf and hard-of-hearing individuals' preferences for wearable and mobile sound awareness technologies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, paper 46. ACM.

9. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. 2018. Datasheets for datasets. arXiv preprint arXiv:1803.09010.
10. Goodman, B., & Flaxman, S. 2017. European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, 38(3), 50-57.
11. Gurari, D., Li, Q., Lin, C., Zhao, Y., Guo, A., Stangl, A., & Bigham, J. P. 2019. VizWiz-Priv: a dataset for recognizing the presence and purpose of private visual information in images taken by blind people. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 939-948.
12. Kacorri, H. 2017. Teachable machines for accessibility. *ACM SIGACCESS Accessibility and Computing*, (119), 10-18.
13. Kacorri, H., Kitani, K. M., Bigham, J. P., & Asakawa, C. 2017. People with visual impairment training personal object recognizers: Feasibility and challenges. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5839-5849. ACM.
14. MacLeod, H., Bennett, C. L., Morris, M. R., & Cutrell, E. 2017. Understanding blind people's experiences with computer-generated captions of social media images. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 5988-5999. ACM.
15. Microsoft. (2019). Seeing AI. Retrieved July 2, 2019 from https://www.microsoft.com/en-us/ai/seeing-ai
16. Peng, Y. H., Hsi, M. W., Taele, et al. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, paper 293. ACM.
17. Profita, H., Albaghli, R., Findlater, L., Jaeger, P., & Kane, S. K. (2016, May). The AT effect: how disability affects the perceived social acceptability of head-mounted display use. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4884-4895. ACM.
18. Ribeiro, M. T., Singh, S., & Guestrin, C. 2016. Why should I trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135-1144. ACM.
19. Sicong, L., Zimu, Z., Junzhao, D., Longfei, S., Han, J., & Wang, X. 2017. Ubiear: Bringing location-independent sound awareness to the hard-of-hearing people with smartphones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(2), 17.
20. US Department of Justice. 2011. ADA Requirements: 1. Service Animals. Retrieved July 1, 2019 from https://www.ada.gov/service_animals_2010.htm
21. Zhao, Y., Szpiro, S., Knighten, J., & Azenkot, S. 2016. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 73-84. ACM.
22. Zhao, Y., Wu, S., Reynolds, L., & Azenkot, S. 2018. A face recognition application for people with visual impairments: understanding use beyond the lab. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, paper 215. ACM.

Social Tensions with Head-Mounted Displays for Accessibility

Steven Goodman
University of Washington
Seattle, WA
smgoodmn@uw.edu

Dhruv Jain
University of Washington
Seattle, WA
djain@cs.washington.edu

Jon Froehlich
University of Washington
Seattle, WA
jonf@cs.uw.edu

Brock Craft
University of Washington
Seattle, WA
bcraft@uw.edu

Leah Findlater
University of Washington
Seattle, WA
leahkf@uw.edu

ABSTRACT

Head-mounted displays (HMDs) can enhance accessibility for people with disabilities, but stigmatization and social misperception may hinder use of these devices in social contexts. In this paper, to motivate and encourage future work, we outline key social challenges in using HMDs for accessibility.

CCS CONCEPTS

• **Human-centered computing** → Accessibility theory, concepts and paradigms; • **Social and professional topics** → People with disabilities.

KEYWORDS

accessibility; head-mounted displays; assistive technology; social perceptions

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.
Proceedings of the 1st Workshop on Challenges Using Head-Mounted Displays in Shared and Social Spaces



Figure 1: Examples of HMDs to support accessibility: (top) real-time captions for people who are deaf and hard of hearing [6] and (bottom) visually enhancing information (e.g., magnifying, increasing contrast) for people with low vision [20].

INTRODUCTION

Head-mounted displays (HMDs) hold tremendous potential to enhance accessibility for people with disabilities, such as providing real-time captioning for people who are deaf or hard of hearing [6–8, 17], augmenting vision for people with low vision [4, 20, 23], allowing hands-free computer access for people with motor impairments [13, 15], and providing vocabulary support for people with language impairments [22]; Figure 1 shows examples. At the same time, the use of HMDs for these and other assistive purposes may give rise to unique issues of social acceptability.

Drawing on our experiences designing and evaluating HMD-based approaches to address real-world accessibility problems [6–8, 13, 18, 20, 22] as well as other work, this workshop paper outlines key social acceptability challenges that will need to be addressed for HMDs to achieve widespread success as an assistive technology. Specifically, we detail: (1) tensions between the evolving societal perceptions of HMDs and concerns of stigma and misperception that can arise with assistive technologies, (2) the need to provide assistive support without undue attentional cost and while maintaining expectations for person-to-person communication, and (3) issues of balancing the sensing needs of many of these HMD-based AT systems with the privacy needs of non-wearers. Through this synthesis, we identify open questions and provide guidance for future researchers in this space.

STIGMA AND SOCIAL ACCEPTABILITY

While public perceptions of wearable devices, particularly wrist-worn and health-related sensors, have improved in the past decade, the attitude toward HMDs specifically is far more tepid—even resistant [10, 11]. Koelle *et al.* [12], for example, found that HMDs are distinct from other wearables because they are perceived to be constantly recording information. Employing HMDs for assistive purposes such as those described in the Introduction may yield further complexities due to the stigmatizing effect of ATs [16]. Misperceptions, for example, that AT users are less capable than non-AT users, can lead to social stress and abandonment of AT devices [9, 19]. How the social perceptions of HMDs and ATs may counter or magnify one another is an open question.

As a first step, Profita *et al.* [18] conducted a large-scale survey on perceptions of HMD use to examine how information provided about a wearer's disability and the device's purpose affected perceptions by third-party observers. HMD use (in this case a Google Glass device) was seen as more socially acceptable when there were visual indicators that the wearer had a disability (*i.e.*, dark glasses and a white cane) than when there were no visual indicators; see Figure 2. Moreover, when observers were told that the HMD use was *assistive* (used to help with transit navigation), there was higher acceptance than for non-assistive use (checking email).

For mainstream HMDs that are being repurposed for assistive uses (as opposed to HMDs specifically designed as AT, *e.g.*, [4]), Profita *et al.* [18]'s findings suggest that communicating the assistive nature of the device directly or indirectly to onlookers may counter some of the negative attitudes towards HMDs.



Figure 2: In Profita et al.'s study [18], HMD use was seen as more socially acceptable when there were (top right and bottom) visual indicators that the wearer had a disability (i.e., wore dark glasses and held a white cane) than when (top left) no visual indicators were present.

This conclusion raises an important tension in that, as discussed above, individuals with disabilities can be sensitive to devices that draw unwanted attention to their disability [19]. How can HMDs for assistive purposes be employed in a way that will allow the wearer to disclose that purpose, but only when desired?

EFFECTIVE COMMUNICATION SUPPORT

HMDs are particularly promising for communication support. For example (Figure 3), Jain *et al.* [6–8] explored HMD-based access to spoken conversation for people who are deaf or hard of hearing, through both real-time captioning on an HMD as well as a system that provided visual indicators of sound location to support speechreading (lipreading) in group conversations. Similarly but for a different domain, Washington *et al.* [21] developed a Google Glass application that provides social cues to children with autism through automatic facial expression recognition. Williams *et al.* [22] designed and evaluated an HMD-based approach to provide private visual and auditory vocabulary prompts for people with aphasia, an acquired language disorder.

These projects have yielded many positive findings, for example, that HMD-based vocabulary prompts for people with aphasia can support independent speech and eye contact with a conversation partner [22]. However, important challenges have also emerged. Providing additional information during a task that is already arduous for some individuals (*i.e.*, spoken conversation) can result in high cognitive effort [22] and attention loss [14]. Some HMD devices (*e.g.*, Microsoft HoloLens) obscure the wearer's eyes and facial expressions, which can interfere with face-to-face communication [6]. HMDs also have the potential to reinforce disparities in who takes on the burden of adapting to their conversation partner's needs. For instance, rather than encouraging hearing people to learn sign language, HCI research tends to focus on helping deaf people communicate via speech, thus placing the burden of adaptation on the deaf person (including some of our own projects, *e.g.*, [6, 8]). Because HMDs provide *private* visual and auditory support, they may reinforce this burden disparity more than a shareable form factor like a smartphone or tablet, where the conversation partner can also consume assistive information if desired. Addressing these design challenges will likely require advances in hardware (*e.g.*, to ensure that the wearer's eyes are visible) and user interaction (*e.g.*, to reduce cognitive load), as well as further study of social issues such as burden.

SENSING, DATA COLLECTION, AND DEVICE CAPABILITIES

HMD systems for accessibility often need to continuously capture, process, and potentially archive data streams from microphones, cameras, and/or other sensors (*e.g.*, to provide real-time captions [7], enhanced visual information [23], or conversational cues [22]). But while extensive information about nearby people and phenomena can enable a wide range of AT solutions, unrestrained data collection raises issues of surveillance, (in)voluntary information sharing, and opportunities for misuse.

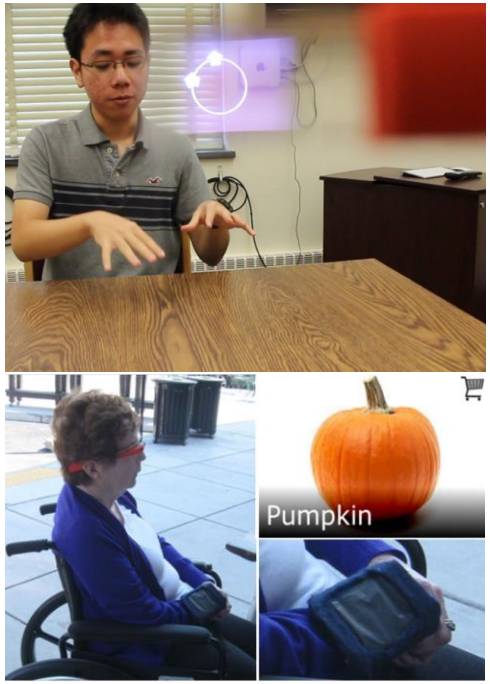


Figure 3: HMDs may be particularly useful for communication support: (top) Jain *et al.* [7] used HMDs to support speechreading in group conversations for deaf and hard of hearing people, (bottom) Williams *et al.* [22] used HMDs to provide visual and auditory vocabulary prompts for people with aphasia.

Recognizing some of these concerns, Ahmed *et al.* [2] investigated the use of HMDs in a workplace environment and found that sighted individuals were more comfortable sharing personal data such as demographics, contact information, and social media accounts with visually impaired colleagues using HMDs as assistive devices than with other sighted colleagues. Additionally, participants were even more comfortable providing this information when they were assured that they could control how their information would be shared. While researchers have begun to investigate solutions for preserving the privacy of observers / bystanders for general HMD use (*e.g.*, using built-in facial obfuscation algorithms for observers [3]), Ahmed *et al.* [2]'s study highlights that privacy concerns and potentially successful solutions may play out differently in an accessibility context.

Individuals with disabilities have also been surveyed about their preferred device capabilities [1, 5, 19]. To provide useful assistive feedback, devices must be capturing information about the world around them. But where and when does this functionality turn off, if ever? Some people with disabilities would greatly benefit from the ability to store and recall information. How can this be done with respect for observer privacy, and how far back should the device recall? Finally, the capabilities of these devices may be abused in more sensitive contexts (bathrooms, meetings, confidential documents, accidental eavesdropping, *etc.*). But if a non-disabled person has access to this information, why shouldn't a disabled person?

A CLOSING NOTE

HMDs for accessibility is a promising research area, and we hope this synthesis of social challenges will guide researchers in building more effective HMD-based AT systems.

ACKNOWLEDGMENTS

This paper was funded in part by the National Science Foundation under grant IIS-1763199.

REFERENCES

[1] Tousif Ahmed, Roberto Hoyle, Patrick Shaffer, Kay Connelly, David Crandall, and Apu Kapadia. 2017. Understanding the Physical Safety, Security, and Privacy Concerns of People with Visual Impairments. *IEEE Internet Comput.* 21, 3 (May 2017), 56–63. https://doi.org/10.1109/MIC.2017.77

[2] Tousif Ahmed, Apu Kapadia, Venkatesh Potluri, and Manohar Swaminathan. 2018. Up to a Limit?: Privacy Concerns of Bystanders and Their Willingness to Share Additional Information with Visually Impaired Users of Assistive Technologies. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3 (September 2018), 89:1--89:27. https://doi.org/10.1145/3264899

[3] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-mediating Technologies. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (CHI '14), ACM, New York, NY, USA, 2377–2386. https://doi.org/10.1145/2556288.2557352

[4] eSight. 2018. Life-changing glasses for the visually impaired. Retrieved from https://www.esighteyewear.com/homex

[5] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon Froehlich, Raja Kushalnagar, and Angela Carey Lin. 2019. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*, ACM Press, New York, New York, USA, 1–13. https://doi.org/10.1145/3290605.3300276

[6] Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. 2018. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems* (DIS '18 Companion), ACM, New York, NY, USA, 7–11. https://doi.org/10.1145/3197391.3205404

[7] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, ACM, New York, NY, USA, 241–250. https://doi.org/10.1145/2702123.2702393

[8] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People Who Are Deaf and Hard of Hearing. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*, ACM, New York, NY, USA, 81–92. https://doi.org/10.1145/3234695.3236362

[9] Shaun K Kane, Chandrika Jayant, Jacob O Wobbrock, and Richard E Ladner. 2009. 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), ACM, New York, NY, USA, 115–122. https://doi.org/10.1145/1639642.1639663

[10] Norene Kelly. 2017. All the World's a Stage: What Makes a Wearable Socially Acceptable. *Interactions* 24, 6 (October 2017), 56–60. https://doi.org/10.1145/3137093

[11] Norene Kelly and Stephen Gilbert. 2016. The WEAR Scale: Developing a Measure of the Social Acceptability of a Wearable Device. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16), ACM, New York, NY, USA, 2864–2871. https://doi.org/10.1145/2851581.2892331

[12] Marion Koelle, Matthias Kranz, and Andreas Möller. 2015. Don'T Look at Me That Way!: Understanding User Attitudes Towards Data Glasses Usage. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (MobileHCI '15), ACM, New York, NY, USA, 362–372. https://doi.org/10.1145/2785830.2785842

[13] Meethu Malu and Leah Findlater. 2015. Personalized, Wearable Control of a Head-mounted Display for Users with Upper Body Motor Impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), ACM, New York, NY, USA, 221–230. https://doi.org/10.1145/2702123.2702188

[14] Gerard McAtamney and Caroline Parker. 2006. An Examination of the Effects of a Wearable Display on Informal Face-to-face Communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '06), ACM, New York, NY, USA, 45–54. https://doi.org/10.1145/1124772.1124780

[15] Roisin McNaney, John Vines, Daniel Roggen, Madeline Balaam, Pengfei Zhang, Ivan Poliakov, and Patrick Olivier. 2014. Exploring the Acceptability of Google Glass As an Everyday Assistive Device for People with Parkinson's. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), ACM, New York, NY, USA, 2551–2554. https://doi.org/10.1145/2556288.2557092

[16] Howard P Parette, Mary Blake Huer, and Marcia Scherer. 2004. Effects of Acculturation on Assistive Technology Service Delivery. *J. Spec. Educ. Technol.* 19, 2 (2004), 31–41. https://doi.org/10.1177/016264340401900203

[17] Yi-Hao Peng, Ming-Wei Hsi, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), ACM, New York, NY, USA, 293:1--293:10. https://doi.org/10.1145/3173574.3173867

[18] Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K Kane. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), ACM, New York, NY, USA, 4884–4895. https://doi.org/10.1145/2858036.2858130

[19] Kristen Shinohara and Jacob O Wobbrock. 2011. In the Shadow of Misperception: Assistive Technology Use and Social Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11), ACM, New York, NY, USA, 705–714. https://doi.org/10.1145/1978942.1979044

[20] Lee Stearns, Leah Findlater, and Jon E Froehlich. 2018. Design of an Augmented Reality Magnification Aid for Low Vision Users. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '18), ACM, New York, NY, USA, 28–39. https://doi.org/10.1145/3234695.3236361

[21] Peter Washington, Catalin Voss, Aaron Kline, Nick Haber, Jena Daniels, Azar Fazel, Titas De, Carl Feinstein, Terry Winograd, and Dennis Wall. 2017. SuperpowerGlass: A Wearable Aid for the At-Home Therapy of Children with Autism. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3 (September 2017), 112:1--112:22. https://doi.org/10.1145/3130977

[22] Kristin Williams, Karyn Moffatt, Denise McCall, and Leah Findlater. 2015. Designing Conversation Cues on a Head-Worn Display to Support Persons with Aphasia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), ACM, New York, NY, USA, 231–240. https://doi.org/10.1145/2702123.2702484

[23] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. ForeSee: A Customizable Head-Mounted Vision Enhancement System for People with Low Vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '15), ACM, New York, NY, USA, 239–249. https://doi.org/10.1145/2700648.2809865

Evaluating Angular Accuracy of Wrist-based Haptic Directional Guidance for Hand Movement

Jonggi Hong[*], Lee Stearns[*], Tony Cheng[*], Jon E. Froehlich[*], David Ross[†], Leah Findlater[‡]

[*]Department of Computer Science
University of Maryland, College Park

[†]Atlanta VA R&D Center of Excellence in Vision and Neurocognitive Rehabilitation (CVNR)

[‡]College of Information Studies
University of Maryland, College Park

ABSTRACT

Haptic guidance for the hand can offer an alternative to visual or audio feedback when those information channels are overloaded or inaccessible due to environmental factors, vision impairments, or hearing loss. We report on a controlled lab experiment to evaluate the impact of directional wrist-based vibro-motor feedback on hand movement, comparing lower-fidelity (4-motor) and higher-fidelity (8-motor) wristbands. Twenty blindfolded participants completed a series of trials, which consisted of interpreting a haptic stimulus and executing a 2D directional movement on a touchscreen. We compare the two conditions in terms of movement error and trial speed, but also analyze the impact of specific directions on performance. Our results show that doubling the number of haptic motors reduces directional movement error but not to the extent expected. We also empirically derive an apparent lower bound in accuracy of ~25° in interpreting and executing on the directional haptic signal.

Keywords: Wearables, haptics, non-visual directional guidance.

Index Terms: H.5.m. Information interfaces and presentation (*e.g.,* HCI)

1 INTRODUCTION

Haptic guidance for the hand can be useful for a range of applications, including virtual reality [26, 33], teaching new motor skills [11, 20], and accessibility for blind users [13, 30]. Such guidance can offer an alternative to visual or audio feedback when those information channels are overloaded or inaccessible due to environmental factors, vision impairments, or hearing loss. For example, a driver may need to interact non-visually with their car's console [4], a firefighter may require fine-grained guidance in a smoky room with no visibility [9], or a blind user may want to accurately trace a line of printed text with a wearable system while listening to text-to-speech output of that text [13, 30].

While researchers have looked at both finger-worn [13, 30] and wrist-worn haptic directional guidance [11, 26, 28, 33], the wrist offers a balance of proximity to the finger, sensitivity [7], surface area, and social acceptability. The larger surface area, for example, allows for a greater number of vibration sources than is possible on the finger. Moreover, emerging smartwatches offer a compelling and timely opportunity to embed haptic feedback in the watchband itself to support a range of non-visual interactions.

Wrist-based haptic feedback, however, has primarily been studied for notifications, such as providing pulse patterns with a single motor (*e.g.,* [24]) or localizing one motor in a grid of motors (*e.g.,* [5, 22]). Fine-grained directional guidance for the hand has received less attention. Weber *et al.* [33] and Sergi *et al.* [28], for example, studied wrist-based haptic guidance for 3D hand movements and a relatively small set of target directions (4 or 6). However, the accuracy with which users can interpret an arbitrary target direction (*e.g.,* our study evaluates 32 directions) and execute a corresponding movement is not yet known—basic information that would impact the design of a range of interactions, from quickly finding a target to tracing a path.

In this paper, we report on a controlled lab experiment to evaluate the impact of wrist-based vibro-motor feedback on hand movement, comparing lower-fidelity (4-motor) and higher-fidelity (8-motor) wristbands. Twenty participants completed a series of trials consisting of interpreting a haptic stimuli and executing a 2D directional movement on a touchscreen. For the results to apply to non-visual interaction scenarios such as situational impairments, participants were blindfolded. We assess movement error and trial speed, but also analyze the impact of specific directions on performance. The contributions include: (i) empirical evidence that doubling the number of haptic motors reduces directional movement error, and that movement error is greater to the upper-left than in other directions; (ii) identification of an empirically derived maximum threshold of ~25° accuracy using our approach; (iii) design considerations for incorporating directional haptic guidance into a smartwatch band—in particular, for most applications, a four-motor wristband may be sufficient.

2 RELATED WORK

Designing effective haptic feedback requires a consideration of factors such as human perception thresholds and the ability to discriminate different haptic patterns; see [6] for a survey. While our focus is on fine-grained directional guidance for hand movements, many projects have looked at directional feedback for directing the user's whole body (*e.g.,* for navigation). These include a few examples with wrist-worn devices with one [2, 3, 18] or four [23] vibro-motors, but more commonly other parts of the body are used, such as the waist or back [8, 17]. In terms of general feedback on the wrist or forearm, Cholewiak *et al.* [7] showed that localization is most precise when the stimulus is close to an anatomical point of reference (*e.g.,* wrist or elbow).

Most closely related to our study is work on haptic directional guidance for the hand or arm. Schätzle *et al.* [26] studied different fidelities of wrist-based translation and rotation feedback, but the study had critical validity issues including no counterbalancing. In follow-up work, Weber *et al.* [33] compared wristband conditions with four or six motors, assessing the user's ability to perceive a signal and rotate or move their arm in one of four or six directions. Providing a verbal direction ("up", "down", "left", or "right") resulted in better performance than the haptic feedback for this relatively simple task. In contrast, we evaluate movement in 32 directions, for which discrete verbal feedback would not be feasible. Sergi *et al.* [28] compared the impact of visual and haptic feedback on reaching accuracy in a virtual reality (visual) environment with a small set of targets. Stanley and

[*]email: {jhong12, lstearns, jonf}@umd.edu
[†]email: ross0128@bellsouth.net
[‡]email: leahkf@umd.edu

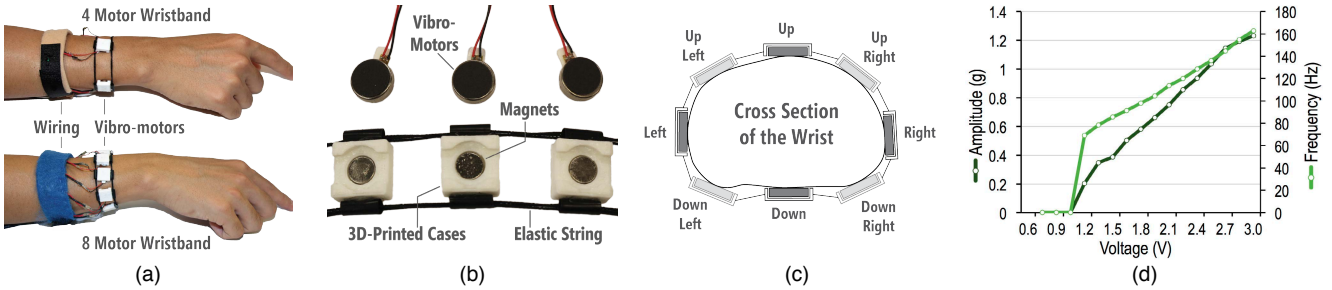


Figure 1: (a) The four- and eight-motor haptic wristband prototypes. **(b)** A close-up of our haptic design; motors were placed in the 3D-printed cases using magnetic attachments, and the wristband faced downward so that the motors directly contacted the skin. **(c)** Motor placement on the right-hand wrist, from wearer's perspective. Our custom-designed wristband could accommodate wrists of various shapes and sizes. Motor placement was adjusted per participant to match this figure. **(d)** Graph of effective voltage *vs.* vibration frequency and amplitude using a wrist-worn motor measured with an oscilloscope and an accelerometer respectively. The x-axis is the analog voltage output on the Arduino Mega, which simulates analog output using PWM. Results were used to verify a linear relationship of voltage input with amplitude and frequency and to empirically determine minimum voltage (1.2V).

Kuchenbecker [29] examined bi-directional rotation guidance on the wrist (guidance to rotate the wrist left or right) by manipulating intensity to indicate the degree of rotation required. Their approach thus provides 1D feedback in contrast to our 2D guidance.

As opposed to these wristband designs, another approach is to arrange motors in a grid or on the back of a watch face rather than along the band. For example, Chen *et al*. [5] compared 3×3 grids of motors on the dorsal and volar sides of the wrist, concluding that only two motors could be reliably distinguished. To indicate cardinal directions (among other information), Yatani and Truong [34] used a cross-layout of five actuators on the back of a phone, while Lee *et al*. [19] used sensory saltation with a 3×3 grid on the back of a smartwatch display. Our prototypes use motors along a wristband, although grids could be explored in the future. Matscheko *et al*. [21] compared the user's ability to discriminate motors placed on the back of a watch face versus on the band, finding higher perception with the band.

Our design relies on a phenomenon called *phantom sensation*, where two vibrotactile actuators placed closely together on the skin create the illusion of a single vibration between the two actuators [1]. In early work, Alles [1] showed that the location of a phantom sensation can be changed by varying the amplitude of the two motors (amplitude inhibition) or offsetting the time at which each motor vibrates (temporal inhibition), but concluded that amplitude variation was most effective. While Alles originally suggested theoretically that it would be better to vary amplitude logarithmically rather than linearly, there is some debate (*e.g.*, [16]), and Seo and Choi [27] recently showed that users were better able to localize the phantom sensation with linear interpolation. For our work, we use linear interpolation.

3 Method

To compare directional guidance with four or eight wrist-mounted vibration motors, we conducted a controlled lab experiment assessing participants' ability to move their finger in a prompted direction. Participants were blindfolded to help simulate a non-visual context and limit distraction. The main hypotheses were that the two wristband conditions would impact accuracy and trial completion time differently, with the eight-motor wristband being particularly accurate.

3.1 Participants

We recruited 20 participants (10 male) aged 19-58 years old (M = 29.7, SD = 10.6) through campus email lists. All but two participants reported daily use of a touchscreen device, and all participants had experience with vibration feedback from a smartphone, video game controller, or other device. Participants were volunteers and compensated $10.

3.2 Apparatus

The custom experimental system consisted of a vibrotactile wristband connected to an Arduino Mega microcontroller, and an Android application running on a Samsung Galaxy Tab 4 (10.1-inch screen, 1280 × 800 resolution). The experimental tasks were presented on the tablet, which communicated with the Arduino via Bluetooth. For each trial, the participant interpreted a vibration and moved their finger correspondingly on the tablet. The touchscreen allowed us to capture precise finger traces in reaction to the stimuli. During the study, the tablet was affixed in a landscape position on a table in front of the participant.

3.2.1 Physical prototype design

The two wristbands, shown in Figure 1a, were identical except for the number of motors (four or eight). We used circular eccentric rotating mass (ERM) disc vibro-motors 10 mm in diameter, with maximum voltage of 5V, maximum frequency of 183Hz, and response time of 100ms.[1] This type of motor has been effectively used to create phantom sensations [25], is inexpensive and ubiquitous, and the flat design means it can be easily integrated into a wristband.

Our custom design addresses three issues: vibration transfer along the band, variation in wrist size, and the non-uniform shape of a wrist. To effectively isolate the motors and limit vibration transfer—an issue in our early designs—we mounted the motors on a band separate from the wiring and housed them magnetically in 3D-printed cases connected only by thin elastic thread (Figure 1b). The band with the wiring connected to the Arduino for communication and power. Because the wrist is not a uniform oval (Figure 1c), placing the motors equidistantly around the band (as in [26, 33]) means that the motors are not necessarily at the position the user expects—for example, the right-most location on the wrist may not be midway between the up and down positions. To address variation in wrist sizes and shapes, our prototype is adjustable. The band with the motors is threaded through the motor cases rather than affixed, which allows the cases to slide (with effort) along the band, and allows the band to be tightened or loosened (based on [14]). The experimenter could thus adjust the band per user.

[1] Adafruit motor disc: http://www.adafruit.com/product/1201

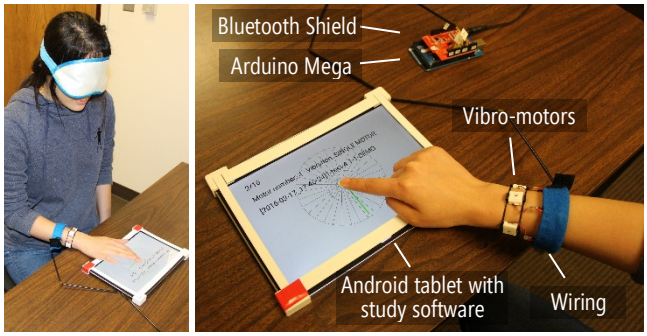


Figure 2: The experimental setup showing the Android tablet, 4-motor wristband, and the Arduino Mega with Bluetooth Shield.

3.2.2 Haptic feedback

Vibration amplitude and frequency were controlled by the Arduino's analogWrite() function, which controls output voltage using pulse-width-modulation (PWM). A higher voltage corresponded to higher vibration amplitude and higher frequency in the ERMs (Figure 1d). ERMs do not provide independent control of amplitude and frequency, but frequency is not a significant factor in accurately locating a phantom vibration [7].

The amplitude range used for the experiment was 0.2–1.2g, while the corresponding frequency range was 69–162 Hz. These ranges were determined through two mechanisms. First, to empirically investigate the physical performance of our ERMs, we used an accelerometer to measure amplitude and frequency at different voltages with an ERM worn on the wrist. As shown in Figure 1d, the relationship was roughly linear above 1.2V for both amplitude and frequency, up to a maximum amplitude of 1.2g and frequency of 162Hz at 3V. Second, to ensure that the full vibro-range used in the experiment would be perceptible by users, we conducted a simple perception test with six participants from our lab. We increased the voltage continuously from 0V and noted the threshold at which the participant first reported feeling a light vibration, then decreased it from 3V until no vibration was felt (again, marking the voltage threshold). To be conservative, we selected the maximum value among all reported thresholds (corresponding to 0.2g and 69Hz) as a perceptible lower bound on our vibration range.

To indicate a target direction, one or two motors vibrated. If the target direction exactly matched a motor's location on the band, the maximum voltage was applied (3.0V) to that single motor. For directions between two motors, voltage was linearly interpolated to produce correspondingly linearly interpolated amplitudes and frequencies within the ranges given above. For example, 45° is exactly between two motors on the 4-motor wristband, so both motors vibrated at the midpoint of the amplitude and frequency ranges, while for 70°, one motor vibrated at higher amplitude and frequency than the other. More formally:

$$V_1 = (\theta_2 - \theta - \theta_{start}) \frac{V_{max} - V_{min}}{\theta_2 - \theta_1 - \theta_{start}} + V_{min},$$

$$V_2 = (\theta - \theta_1 - \theta_{start}) * \frac{V_{max} - V_{min}}{\theta_2 - \theta_1 - \theta_{start}} + V_{min}$$

where V_1 and V_2 are the voltages applied to the two neighboring motors, $V_{max} = 3.0$, $V_{min} = 1.2$, θ_1 and θ_2 indicate the motor placements, and θ is the target direction ($\theta_2 > \theta > \theta_1$). To support 32 discrete directions for our experimental task (11.25° intervals; Section 3.3), interpolation only occurred if θ is >= 11.25° from any motor; θ_{start} is thus a constant equal to 11.25°.

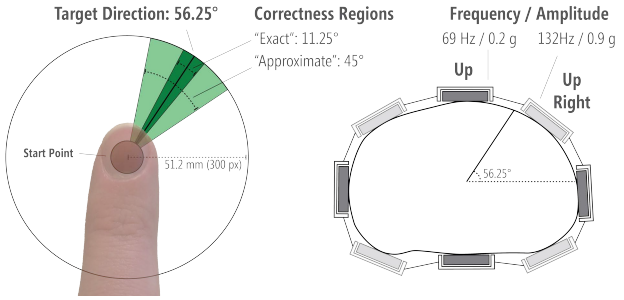


Figure 3: Example trial with target angle of 56.25°, showing start location, "exact" and "approximate" correctness regions, and relative frequency and amplitude of the two neighboring motors used to interpolate the target angle.

3.3 Procedure

The study began with a background survey. The two experimental conditions (4-motors or 8-motors) were then presented in counterbalanced order. For each, the experimenter placed the wristband on the participant, adjusted the size to be tight yet comfortable, positioned the vibration motors as shown in Figure 1c, and moved the tablet perpendicular to the participant's arm. We did not directly control the participant's arm placement, but asked them to use the same posture throughout. Participants were then blindfolded to ensure non-visual interaction. See Figure 2 for pictures of the experimental setup.

The task included 32 uniformly distributed target directions (11.25° intervals). The number of directions (32) was chosen both because it is divisible by 8—thus covering all cardinal and intercardinal directions—and because it offers greater precision than experienced members of our research team could achieve when using the wristbands—thus likely allowing us to identify an upper bound on human performance with this task.

For each trial, the participant placed their finger on a tactile marker (a small, thin sticker) in the middle of the screen. After one second, a chime played and the wristband began vibrating in a target direction. The participant moved their finger in that direction until the vibration stopped and audio feedback played, which happened when the finger was more than 51.2 mm (300 px) from the start. The vibration stimulus was thus applied for the full duration of the trial, which was about two seconds in our study. As shown in Figure 3, we provided two levels of success feedback so that participants could aim for precision but not be discouraged: "exact" success was within an 11.25° interval centered at the target angle (a chime followed by the word 'Perfect!"), while "approximate" success was within a 45° interval (only the chime). Otherwise, the trial was an error (beep sound).

Participants completed 16 practice trials: eight directions at 45° counterclockwise intervals, and eight more directions randomly selected from the full 32 (but the same for all participants). During practice, participants could repeat error trials once. The test task consisted of 96 trials: three repetitions of each of the 32 directions presented in randomized order per participant. Participants were asked to complete the task quickly and accurately. Afterward, participants compared the two conditions (4-motors and 8-motors) on ease of use, accuracy, and preference.

3.4 Experiment Design and Hypotheses

We used a within-subjects design with a single factor of *number of motors*: four or eight motors. Order of presentation was fully counterbalanced and participants were randomly assigned to orders. Our hypotheses were:

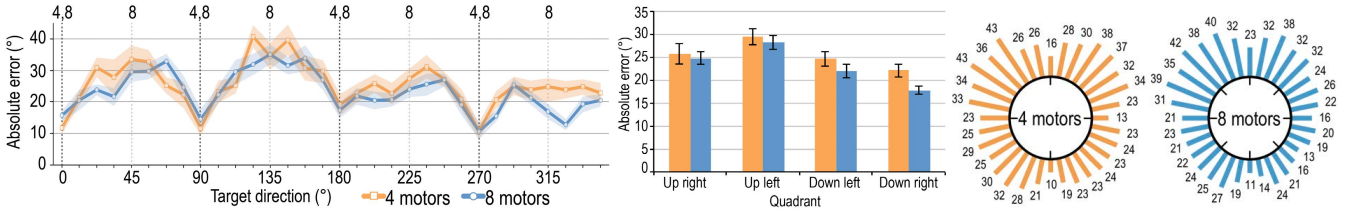


Figure 4: (a) Average absolute error per direction, showing lower error in cardinal directions but a less clear trend for intercardinal ones, even for the 8-motor wristband. The numbers at the top of the graph (4, 8) indicate directions corresponding to physical motor placement (4- or 8-motor bands).Shading indicates standard error. **(b)** Absolute error by quadrant. Error bars indicate standard error. **(c)** Average absolute error by angle, showing highest errors in top-left quadrant. *N* = 18 for all three figures (two participants were extreme outliers and not included).

- *H1: The 8-motor condition will result in lower absolute error than the 4-motor condition*. Because eight motors provide twice the fidelity, participants should be able to more precisely interpret the vibration than with four motors.
- *H2: The 8-motor and 4-motor conditions will impact trial completion time differently*. For this measure, which encompasses both the time to perceive the stimulus and execute a motion, we did not have a directional hypothesis. Conceivably, the 4-motor condition could be faster because it is simpler and should be perceptually easier for at least the cardinal directions, or the 8-motor condition could be faster because of the additional intercardinal information.

3.5 Data and Analysis

All touchscreen interactions during the trials were logged. The primary measures were *absolute error*, defined as the absolute angular difference between the target angle and movement angle, and *trial completion time*, defined as the time per trial measured from the start of directional vibration. We also examined *signed error* as a secondary measure to assess any systematic clockwise or counterclockwise bias to the movements.

In total, participants completed 1920 trials: 20 participants × 96 trials. Two participants were extreme outliers on the primary measure of absolute error for at least one condition using the inter-quartile range method [32]; they are thus excluded from analysis. One of these participants was the oldest of the set, while the second encountered substantial confusion in recognizing vibrations on the top of his wrist during the first condition. In general, we used parametric tests when the data met applicable assumptions, and non-parametric tests otherwise. For our primary measures specifically, the absolute error measure did not violate the normality assumption (*i.e.,* Shapiro-Wilk tests for the 4-motor and 8-motor conditions were not significant). Thus, since hypothesis H1 was directional, we used a one-tailed paired t-test for this measure (the only one-tailed test used in the entire analysis). Trial completion time (speed) did violate normality in the 4-motor condition (Shapiro-Wilk test: $W = 0.88$, $p = .029$), so we used a Wilcoxon signed rank test in this case.

4 Results

We compare the two conditions in terms of error and speed, and provide secondary, more in-depth directional analyses.

4.1 Movement Error and Trial Success Rates

For both conditions the absolute error was about twice the interval between the target directions. As expected, participants were more accurate with the 8-motor condition than with the 4-motor condition, with an average absolute movement error of 23.2° (*SD* = 3.47) compared to 25.4° (*SD* = 4.6). This difference was statistically significant with a small-to-medium effect size, supporting hypothesis H1 (t_{17}=–1.95, $p = .034$, $d = 0.46$). This difference in absolute error, however, did not translate to significant differences in task success. In both conditions, participants had difficulty completing trials exactly correct (*i.e.*, within the 11.25° interval). As shown in Table 1, roughly 16% of trials were completed exactly correct regardless of condition, while just over half were completed approximately correct.

Threshold	Exact	Approximate
4 motors	15.9% (*SD*=5.2%)	54.3% (*SD*=8.2%)
8 motors	16.3% (*SD*=2.9%)	56.4% (*SD*=6.5%)

Table 1: Average number of exactly or approximately correct trials. A movement within an 11.25° interval of the prompted direction was considered *exactly* correct, while movement with a 45° interval was considered *approximately* correct.

Finally, to understand whether there was a systematic bias in the direction of movement error either clockwise (+) or counterclockwise (-), we analyzed signed error. The average signed error was close to zero for both conditions: –1.3° (*SD* = 7.6) with four motors and 0.8° (*SD* = 7.9) with eight motors. This suggests that error in either direction (clockwise or counterclockwise) tended to cancel out on average. A paired t-test comparing the two conditions was not significant.

4.2 Trial Completion Time

The average completion time per trial was roughly two seconds: 2268.5 ms (*SD* = 767.2) with four motors and 2058.2 ms (*SD* = 664.6) with eight motors. The difference in average completion times between conditions was not significant with a Wilcoxon signed ranks test ($Z = -.89$, $p = .372$). The majority of this time consisted of perceiving the stimulus and planning the motor movement—the start of finger movement only began on average after 1707.8 ms (*SD* = 506.5) with 4 motors and 1557.9 ms (*SD* = 466.9) with 8 motors. Hypothesis H2, that there would be a speed difference between the two conditions, is not supported.

4.3 Detailed Directional Analysis

As a secondary analysis, we explored the extent to which the target angle impacted performance; Figure 4. We expected that the 8-motor condition would outperform the 4-motor condition for intercardinal directions (45°, 135°, 225°, and 315°), but as shown in Figure 4a, this pattern was not as clear as predicted. The average absolute error for these directions was 30.0° (*SD* = 8.7) with four motors versus 26.4° (*SD* = 7.0) with eight motors. This difference was only a trend-level result, suggesting that a larger sample size may be needed (paired t-test: $t_{17} = -1.87$, $p = .08$). Absolute error was lower with the cardinal directions, at 13.3° (*SD* = 6.0) for four motors and 14.6° (*SD* = 5.1) for eight motors,

but the difference between the two wristband conditions was not significant (paired t-test: $t_{17} = 0.67$, $p = .510$).

More interesting is a breakdown of error by quadrant. As Figures 4b and 4c show, error tended to be higher in the top-left quadrant than in other directions. A two-way repeated measures ANOVA (number of motors × quadrant) was conducted for absolute error.[2] A significant main effect showed that the quadrants impacted the error ($F_{2.0,41.1} = 10.38$, $p < .001$, $\eta^2 = .38$). Posthoc comparisons with a Bonferroni adjustment showed that the upper-left quadrant was significantly worse than the two downward quadrants (at $p < .05$). The main effect of number of motors was only a trend ($F_{1,41.1} = 3.81$, $p = .068$, $\eta^2 = .18$), and the interaction effect was not significant.

4.4 Subjective Response

While no open-ended comments were collected, participants selected their preferred wristband in terms of ease of use, accuracy and overall preference. For accuracy, 13 participants chose the 8-motor wristband, suggesting they perceived value in the extra information it provided over the 4-motor wristband. Participants were divided, however, in terms of overall preference (9 *vs*. 9) and ease of use (10 votes for the 4-motor wristband *vs*. 8 votes for the 8-motor one).

5 Discussion and Conclusion

Our results confirm one of our two hypotheses, showing that doubling the number of haptic motors from four to eight increases accuracy of wrist-based directional guidance of the hand; however, speed was not impacted either way. Practically speaking, whether to include four or more motors for directional guidance will depend on the accuracy needed for the task. For example, tracing a shape or route would require higher accuracy than simply finding an object with the hand; the latter task could even be completed successfully even if the hand only moved in cardinal or intercardinal directions. Thus, the added cost and weight of doubling the number of motors may not be worth what was only an extra two degrees of accuracy in our study. Thus, we recommend that in most cases only four motors be used.

We also empirically identified a potential upper bound on accuracy with our approach, about ~23-25°. This accuracy is high enough to show that the frequency interpolation between adjacent motors was effective at least in the 4-motor condition. But, it also suggests that other human perceptual or motor factors are at play. While our study purposely did not isolate these two factors—we were interested in holistically evaluating the impact of haptic guidance on hand movement—it is an important area for future work. For example, a future study could present a visual array of 32 uniformly distributed target directions (at 11.25° intervals) and ask participants to select the target corresponding to the perceived vibro-stimulus; this would limit the impact of human motor factors.

Detailed analyses of the impact of target direction on error also showed that movements in the upper-left quadrant were less accurate than other quadrants, particularly the bottom quadrants. This pattern may be due to all participants completing the task with their right hand, and the upper-left quadrant requires more gross motor movement than the bottom quadrants. It is also possible that there is lower acuity at that location on the wrist due to factors such as bone structure or nerve density. Further work, including with left-handed users, is needed.

[2] Greenhouse-Geisser adjustment used, so degrees of freedom are fractional.

Our design included ERM vibro-motors embedded in a wristband, with the motivation that such a solution could be incorporated at low cost into a smartwatch. Other vibro-motor types, however, offer different characteristics that may be worth considering. For example, while the response time of our motors (~100ms) was much less than the overall time before participants initiated movement (> 1.5s), a design similar to ours but with linear resonant actuators (LRAs) could be useful to explore. LRAs offer shorter response time and allow for independent control of amplitude and frequency. Also, we applied linear interpolation to manipulate the perceived location of vibration, but since there is some debate about whether linear or logarithmic interpolation is better [1, 16], future work should compare the two. Another alternative is to use a grid of actuators perhaps placed on the back of a watch face to indicate at least coarse-grained direction (*e.g.,* [15, 19, 22]); such an approach could be compared to our band design. In addition, while we used a two-band design to reduce vibration transfer and adjust the motors per participant, we envision watchband designs with embedded motors and wiring. More work is needed to investigate vibration transfer using other form factors and band materials.

Future work should also investigate how our identified accuracy limit affects performance for more realistic tasks with continuously updated directional guidance (*e.g.,* finding a target, or tracing a path). While our current study did not include blind users, one long-term goal of our work is to extend these findings to include users with visual impairments and to investigate the use of haptic feedback for tasks such as physically exploring a printed page to read text or to gain spatial layout information (*e.g.,* [10, 31]). In these cases, the haptic stimulus should dynamically update to guide the user across an object or towards a target. In our study, vibrations were constant and did not update based on the user's movement (*e.g.,* to help correct an erroneous movement). This decision was purposeful so that our task would be more fundamental, but a more dynamic approach would likely result in higher accuracy. Of course, in any real-world task, potential sensory adaptation to the haptic vibration would need to be considered [12].

References

[1] Alles, D.S. 1970. Information Transmission by Phantom Sensations. *IEEE Transactions on Man-Machine Systems*. 11, 1 (1970), 85–91.

[2] Bosman, S., Groenendaal, B., Findlater, J., Visser, T., de Graaf, M. and Markopoulos, P. 2003. GentleGuide: An exploration of haptic output for indoors pedestrian guidance. *Human-Computer Interaction with Mobile Devices and Services* (2003), 358–362.

[3] Brock, A., Kammoun, S., Macé, M. and Jouffrais, C. 2014. Using wrist vibrations to guide hand movement and whole body navigation. *i-com*. 13, 3 (Jan. 2014), 19–28.

[4] Burnett, G.E. and Porter, J.M. 2011. Ubiquitous computing within cars: designing controls for non-visual use. *International Journal of Human-Computer Studies*. 55, 4 (2011), 521–531.

[5] Chen, H.Y., Santos, J., Graves, M., Kim, K. and Tan, H.Z. 2008. Tactor localization at the wrist. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 5024 LNCS, (2008), 209–218.

[6] Choi, S. and Kuchenbecker, K.J. 2013. Vibrotactile display: Perception, technology, and applications. *Proceedings of the IEEE*. 101, 9 (2013), 2093–2104.

[7] Cholewiak, R.W. and Collins, A.A. 2003. Vibrotactile localization

on the arm: effects of place, space, and age. *Perception & Psychophysics*. 65, 7 (2003), 1058–1077.

[8] Cosgun, A., Sisbot, E.A. and Christensen, H.I. 2014. Evaluation of rotational and directional vibration patterns on a tactile belt for guiding visually impaired people. *IEEE Haptics Symposium, HAPTICS*. (2014), 367–370.

[9] Dyrks, T., Denef, S. and Ramirez, L. 2008. An empirical study of firefighting sensemaking practices to inform the design of ubicomp technology. *Sensemaking Workshop of ….* (2008), 1–5.

[10] Findlater, L., Stearns, L., Du, R., Oh, U., Ross, D., Chellappa, R. and Froehlich, J.E. 2015. Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (2015), 383–384.

[11] Guo, W., Ni, W., Chen, I.M., Ding, Z.Q. and Yeo, S.H. 2009. Intuitive vibro-tactile feedback for human body movement guidance. *2009 IEEE International Conference on Robotics and Biomimetics, ROBIO 2009*. (2009), 135–140.

[12] Hahn, J.F. 1966. Vibrotactile adaptation and recovery measured by two methods. *Journal of Experimental Psychology*. 71, 5 (1966), 655–658.

[13] Horvath, S., Galeotti, J., Wu, B., Klatzky, R., Siegel, M. and Stetten, G. 2014. FingerSight: Fingertip Haptic Sensing of the Visual Environment. *IEEE Journal of Translational Engineering in Health and Medicine*. 2, (2014), 9 pages.

[14] How to Make a Sliding Knot (single knot) - jewelry making tutorial: 2012. *https://youtu.be/t6Qy5WdSq74*.

[15] Israr, A. and Poupyrev, I. 2011. Control space of apparent haptic motion. *2011 IEEE World Haptics Conference, WHC 2011*. (2011), 457–462.

[16] Israr, A. and Poupyrev, I. 2011. Tactile Brush : Drawing on Skin with a Tactile Grid Display. *Proceedings of CHI'11*. (2011), 2019–2028.

[17] Jones, L.A., Kunkel, J. and Piateski, E. 2009. Vibrotactile pattern recognition on the arm and back. *Perception*. 38, 1 (2009), 52–68.

[18] Kammoun, S., Jouffrais, C., Guerreiro, T., Nicolau, H. and Jorge, J. 2012. Guiding Blind People with Haptic Feedback. *Pervasive 2012 Workshop on Frontiers in Accessibility for Pervasive Computing* (Jan. 2012).

[19] Lee, J., Han, J. and Lee, G. 2015. Investigating the Information Transfer Efficiency of a 3x3 Watch-back Tactile Display. (2015), 1229–1232.

[20] Lieberman, J. and Breazeal, C. 2007. TIKL: Development of a wearable vibrotactile feedback suit for accelerated human motor learning. *Proceedings - IEEE International Conference on Robotics and Automation*. 23, 5 (2007), 4001–4006.

[21] Matscheko, M., Ferscha, A., Riener, A. and Lehner, M. 2010. Tactor placement in wrist worn wearables. *Proceedings - International Symposium on Wearable Computers, ISWC*. (2010), 1–8.

[22] Oakley, I., Kim, Y., Lee, J. and Ryu, J. 2006. Determining the feasibility of forearm mounted vibrotactile displays. *Proceedings - IEEE Virtual Reality*. 2006, (2006), 27–34.

[23] Panëels, S., Brunet, L. and Strachan, S. 2013. Strike a Pose: Directional Cueing on the Wrist and the Effect of Orientation. *Haptic and Audio Interaction Design*. I. Oakley and S. Brewster, eds. Springer Berlin Heidelberg. 117–126.

[24] Pasquero, J., Stobbe, S.J. and Stonehouse, N. 2011. A haptic wristwatch for eyes-free interactions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)* (New York, New York, USA, 2011), 3257–3266.

[25] Richter, H., Blaha, B. and Wiethoff, A. 2011. Tactile feedback without a big fuss: simple actuators for high-resolution phantom sensations. *Proceedings of the 13th International Conference on Ubiquitous Computing - UbiComp '11*. (2011), 85–88.

[26] Schaetzle, S.. H.T.. P.C. and H.G. 2006. Evaluation of Vibrotactile Feedback to the Human Arm. *EuroHaptics 2006*. (2006), 557–560.

[27] Seo, J. and Choi, S. 2010. Initial study for creating linearly moving vibrotactile sensation on mobile device. *2010 IEEE Haptics Symposium, HAPTICS 2010*. (2010), 67–70.

[28] Sergi, F., Accoto, D., Campolo, D. and Guglielmelli, E. 2008. Forearm orientation guidance with a vibrotactile feedback bracelet: On the directionality of tactile motor communication. *Proceedings of the 2nd Biennial IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, BioRob 2008*. (2008), 433–438.

[29] Stanley, A. a. and Kuchenbecker, K.J. 2012. Evaluation of tactile feedback methods for wrist rotation guidance. *IEEE Transactions on Haptics*. 5, 3 (2012), 240–251.

[30] Stearns, L., Du, R., Oh, U., Wang, Y., Chellappa, R., Findlater, L. and Froehlich, J.E. 2014. The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind. *Workshop on Assistive Computer Vision and Robotics (ACVR'14) in Conjunction with the European Conference on Computer Vision (ECCV'14)*. (2014), 615–631.

[31] Stearns, L., Du, R., Oh, U., Wang, Y., Chellappa, R., Findlater, L. and Froehlich, J.E. 2014. The design and preliminary evaluation of a finger-mounted camera and feedback system to enable reading of printed text for the blind. *Workshop on Assistive Computer Vision and Robotics (ACVR'14) in Conjunction with the European Conference on Computer Vision (ECCV'14)* (Zurich, Switzerland, 2014), 615–631.

[32] Upton, G. and Cook, I. 1996. *Understanding statistics*. Oxford University Press.

[33] Weber, B., Schätzle, S., Hulin, T., Preusche, C. and Deml, B. 2011. Evaluation of a vibrotactile feedback device for spatial guidance. *2011 IEEE World Haptics Conference, WHC 2011*. (2011), 349–354.

[34] Yatani, K. and Truong, K.N. 2009. SemFeel: a user interface with semantic tactile feedback for mobile touch-screen devices. *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, NY, USA, 2009), 111–120.

Evaluating Wrist-Based Haptic Feedback for Non-Visual Target Finding and Path Tracing on a 2D Surface

Jonggi Hong[1], Alisha Pradhan[2], Jon E. Froehlich[1,3], Leah Findlater[2,4]

[1]Department of Computer Science
[2]College of Information Studies
University of Maryland, College Park

[3]Computer Science and Engineering
[4]Human Centered Design and Engineering
University of Washington

{ jhong12, alisha93, jonf, leahkf }@umd.edu

ABSTRACT

Precisely guiding a blind person's hand can be useful for a range of applications from tracing printed text to learning and understanding shapes and gestures. In this paper, we evaluate wrist-worn haptics as a directional hand guide. We implemented and evaluated the following haptic wristband variations: (1) *four* versus *eight* vibromotor designs; (2) vibration from only a *single* motor at a time versus from two adjacent motors using *interpolation*. To evaluate our designs, we conducted two studies: Study 1 (N=13, 2 blind) showed that participants could non-visually find targets and trace paths more quickly and accurately with single-motor feedback than with interpolated feedback, particularly when only four motors were used. Study 2 (N=14 blind or visually impaired participants) found that single-motor feedback with four motors was faster, more accurate, and most preferred compared to similar feedback with eight motors. We derive implications for the design of wrist-worn directional haptic feedback and discuss future work.

CCS Concepts

Human-centered computing→Accessibility→Empirical Studies in Accessibility

Keywords

Haptic feedback; haptic wristband; blind user; wearable computing; accessibility.

1. INTRODUCTION

Directional hand guidance can be useful for a range of everyday activities for people who are blind and visually impaired (VI), such as physically tracing printed text to hear text-to-speech output [36, 37], learning handwriting [4] and touchscreen gestures [26], and understanding shapes [13]. To provide this guidance, researchers have explored both audio feedback, such as sonification and verbal guidance [26], as well as haptics, such as finger-mounted vibromotors [36, 37] or force feedback [13].

Smartwatches present new opportunities for directional hand guidance that is proximal to the hand yet embedded unobtrusively into an existing general-purpose device. For example, haptic motors around the wrist, as shown in Figure 1, may be useful for guiding a blind user's hand on a 2D surface. However, how to design this guidance is an open question. While placing a larger number of haptic actuators around the wrist may offer more precise feedback, it also increases the physical weight and complexity of the device. At the same time, fewer motors may still be useful, particularly if it is possible to create *phantom sensations*, where two closely-placed motors vibrating at once create the perception of a single vibration located between the two [2].

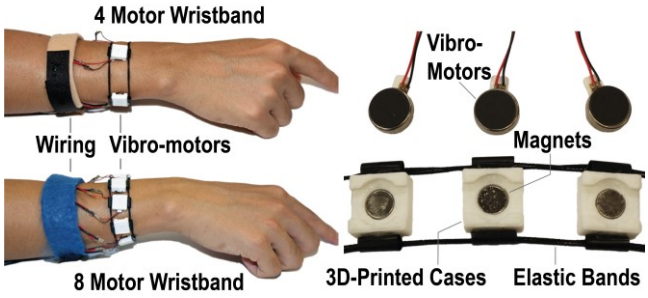


Figure 1. Wristbands with four and eight motors. Image adapted from our prior work [18], which used the same hardware and wristband configuration.

Wrist-worn haptic feedback has been primarily studied with sighted users and, even then, in the context of notifications such as the user's ability to recognize different pulse patterns (*e.g.*, [29]). For blind users, one [7, 8, 20] or four vibromotors [27] have been used on the wrist to support navigation (*i.e.*, directing the user's whole body) but not for precise hand guidance, which is our focus. A small number of studies with sighted users have focused on wrist-worn haptics for directional hand guidance [1, 35, 40]; however, [1, 35] did not compare multiple haptic designs so conclusions about efficacy cannot be drawn. Two exceptions come from Weber *et al.* [40] and Hong *et al.* [18]. Weber *et al.* [40] compared wristbands with four or six haptic motors to verbal instructions (up/left/right/down) to guide sighted users in moving and rotating their arm. However, they did not find statistically significant differences between the haptic designs, and their focus was on guiding hand movement in free space rather than on a 2D surface. Hong *et al.* [18] showed that an 8-motor wristband resulted in small accuracy benefits over a 4-motor design but their experimental task was relatively simple (making a directional swipe) and their study did not include blind users, who may perform differently due to differences in tactile perception (*e.g.*, [16]).

To study directional wrist-worn haptic feedback for blind users and to compare the effectiveness of different feedback designs, we implemented and evaluated the following haptic wristband variations: (1) *four* versus *eight* vibromotor designs; (2) vibration from only a *single* motor at a time versus from two adjacent motors, with the intention of creating an *interpolated*, more precise phantom sensation (vibration) between the two. In theory, eight

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ASSETS '17, October 29-November 1, 2017, Baltimore, MD, USA
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-4926-0/17/10…$15.00
https://doi.org/10.1145/3132525.3132538

vibromotors and the use of phantom sensations should provide more precise feedback than the other options but requires more hardware and could also be more cognitively or physically taxing. We conducted two controlled lab experiments: the first study included 11 sighted and 2 blind participants and was aimed at paring down the number of experimental conditions before conducting a second study with 14 blind and VI participants.

Study 1 showed that participants could non-visually find targets and trace paths more quickly and accurately with single-motor feedback than with interpolated feedback, particularly when only four motors were used. Study 2 (with blind participants) eliminated the interpolated feedback conditions and found that single-motor feedback with four motors was faster, more accurate, and most preferred compared to similar feedback with eight motors. In summary, this paper contributes empirical evidence that wrist-worn haptic feedback with a single motor vibrating at once performs better than interpolated feedback, and that embedding four motors in the wristband performs better than eight motors. We also provide design implications for non-visual wrist-worn haptic guidance. Our results have implications for future smartwatch haptic wristband designs.

2. RELATED WORK

We highlight assistive uses of haptic feedback as well as wrist and hand-worn haptic feedback devices.

2.1 Haptic Feedback for Assistive Applications

Haptic feedback has been used for a range of assistive applications for users with visual impairments. Most effort has focused on sensory substitution of visual information (*e.g.*, [31]) and on navigation (*e.g.*, [15]) but other areas have also been investigated, such as hand-worn dynamic braille for deaf-blind users [3, 17]. Sensory substitution maps visual information to alternative channels such as audio or haptics. A well-known example is *BrainPort*, which translates images from a glasses-worn camera into electrical stimulations on the tongue [31]. Another example is *Optacon*, which translates black and white areas on a printed page to dynamic tactile output for the user to touch [33]. Haptic wristbands and gloves have also been used to convey the color of objects under the finger [9, 34].

In this paper, we explore the use of haptics for directional guidance. Here, most work has focused on whole-body navigational support. An early example comes from Ertan *et al.* [15], who employed a 4x4 grid of vibrotactile actuators on the back of a vest to indicate turn information. Similarly, Scheggi *et al.* [32] used wrist-worn devices with two haptic actuators that were controlled by a remote observer to help blind users avoid obstacles while walking. Mann *et al.* [24] employed a haptic helmet combined with a depth camera to help blind users avoid collisions, although no user evaluation was reported. Cosgun *et al.* [12] evaluated a haptic belt with eight motors to provide directional guidance in discrete, 45° increments. They showed that users could recognize which of the eight motors was vibrating with 55-97% accuracy, depending on the motor. Though these studies showed that haptic feedback devices in different form factors are feasible to use for whole-body navigation, the design of wrist-worn haptic feedback for precise hand guidance has not been investigated with blind users.

2.2 Non-visual Directional Hand Guidance

For non-visual directional hand guidance specifically, prior work has explored both audio and haptic channels. For example, Oh *et al.* [26] used verbal feedback and sonification to teach blind users how to make touchscreen gestures. Audio feedback has also been used to help users reach targets in their peripersonal space [28, 41]. Directional audio guidance, however, has the drawback of interfering with other uses of the audio channel, such as screenreader or text-to-speech output [37]. Haptic feedback addresses this issue. Kim *et al.* [21] used four vibromotors mounted on a handheld smartphone to guide users with visual impairments to find targets on a large display. A user study showed that speech plus haptic feedback was faster than speech alone in helping users find targets. Shilkrot *et al.* [36] and Stearns *et al.* [37] both employed haptic actuators mounted on the finger and audio feedback to provide up/down (1D) finger guidance for a blind user tracing along a line of printed text. The latter conducted a controlled experiment with 19 blind participants to compare haptic and audio feedback for this reading task, finding that there were tradeoffs in accuracy and user preference between the two [37]. In contrast to these studies, our paper focuses on using haptic feedback around the wrist for precise 2D directional guidance for blind users.

Wrist-worn directional haptic guidance has been investigated to a greater extent with sighted users. In addition to the Weber *et al.* [40] example mentioned in the introduction, another study that is closely related to ours evaluated the accuracy with which users can perceive four or eight vibromotors placed around the wrist and move their hand in response [18]. In a user study, they vibrated one or two motors at a time to create a phantom sensation that conveyed direction more precisely than the 90° or 45° angular intervals between the four or eight motors. Users were asked users to move their finger in the direction they perceived. The wristband with eight motors was more accurate than the wristband with four motors, though the angular error in movement ranged from ~23–25° for both wristbands. In comparison to that study, our paper focuses on more complex and realistic tasks (target finding and path tracing), includes blind and visually impaired users, and explores a wider range of haptic feedback designs.

Other haptic studies with sighted users have investigated using a grid of motors on one side of the wrist to convey direction instead of placing motors around a band. For example, Lee *et al.* showed that the accuracy of perceiving direction with the grid of motors on the back of a smartwatch is more accurate with sensory saltation (an illusion of stimulus movement) than without it [22]. Another study empirically showed that the accuracy of localizing a vibration within a 3x3 grid layout was ~22–76% depending on the motor [10]. Finally, Matscheko *et al.* [25] compared a grid layout to a wristband layout for directional guidance while users are engaged in tasks that occupy the visual and audio sensory channels. They found the band layout to be more accurate, which supports our decision to use a design with motors around the band rather than in a grid.

3. DESIGN OF HAPTIC GUIDANCE

To evaluate the performance of different haptic feedback designs on 2D hand movement, we implemented two haptic wristbands: one with four motors and the other with eight motors. They employ the same hardware design as in [18] but with different software.

3.1 Hardware

The wristbands, shown in Figures 1 and 2, are identical except for the number of motors. We use eccentric rotating mass (ERM) disc vibromotors 10mm in diameter. [1] ERMs are inexpensive and ubiquitous, their flat design can be easily integrated into a wristband, and they have been used to create phantom sensations

[1] https://www.adafruit.com/product/1201

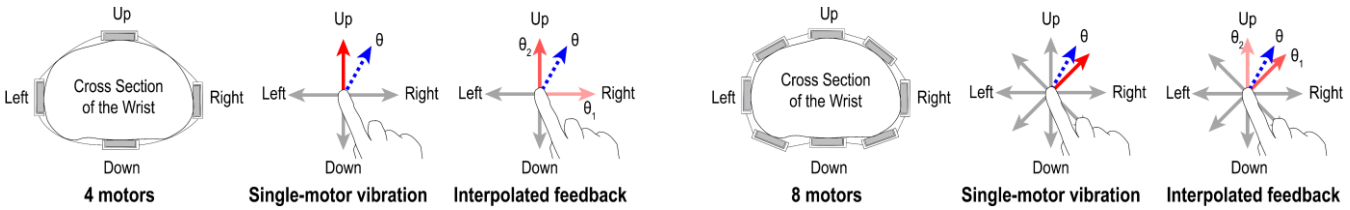


Figure 2. The position of motors around the wrist and an example of haptic feedback of 4 or 8 motors with single motor vibration or interpolation. In this figure, θ is the prompted direction (blue dotted arrow) and red and light red arrows are directions of vibration.

[30]. Additionally, ERMs performed better for our user tasks when compared to a similar wristband design with linear resonant actuators (LRAs) when tested with three internal pilot participants. When mounted on the wristband and worn by a user, we measured the amplitude of the ERMs at 0-1.2g and the frequency at 0-162Hz; Figure 3 shows the measured frequency and amplitude in response to applied voltage. The response time to drive the motor to full vibration from the complete idle state was 100ms.

As shown in Figure 1, our custom wristband design addresses three issues: vibration transfer along the band, variation in wrist size, and the non-uniform shape of a wrist. To isolate the motors and limit vibration transfer, we mounted the motors on a band separate from the wiring and housed them magnetically in 3D-printed cases connected only by thin elastic thread. The band with the wiring connected to an Arduino Mega for communication and power. Because the wrist is not a uniform oval, placing the motors equidistantly around the band (as in [40]) means that the motors are not necessarily at the position the user expects—for example, the right-most location on the wrist may not be midway between the up and down positions. To address variation in wrist sizes and shapes, our prototype is adjustable. The band with the motors is threaded through the motor cases rather than affixed, which allows the cases to slide (with effort) along the band, and allows the band to be tightened or loosened (based on [19]). The experimenter could thus adjust the band per user. The Arduino Mega was in turn connected to a Samsung Galaxy Tab 4 tablet via an on-the-go (OTG) cable. The tablet communicated serially with the Arduino Mega, sending a signal at a frequency of 92 Hz to update the vibration.

3.2 Feedback

We implemented two different vibration types: *single-motor* vibration, which is simpler and thus may be easier to perceive, and *interpolated* vibration, which should offer more precise directional information. To mitigate desensitization that can occur when haptic feedback is continuously applied at a single location [5], our feedback designs provide only *corrective* guidance. That is, when the user's finger moves off course, vibration guides them back to the intended direction (corrects the finger movement). When the finger movement is on course, no vibration occurs. This design was selected among several options based on extensive piloting within our research lab, including options that provided continuous vibration such as dropping to low intensity rather than no vibration.

3.2.1 Single Motor Vibration

For this feedback type, only one motor vibrates at a time, and the location of that motor maps to a direction on a 2D plane (Figure 2). Directions are in 90° increments for the 4-motor wristband and 45° increments for the 8-motor wristband. For example, vibrating the leftmost motor on the wrist indicates that the finger should move to the left on the 2D surface. To convey directions *between* two motors, the single closest motor vibrates (*e.g.*, for a prompted direction of 15°, the rightmost motor, at 0°, vibrates).

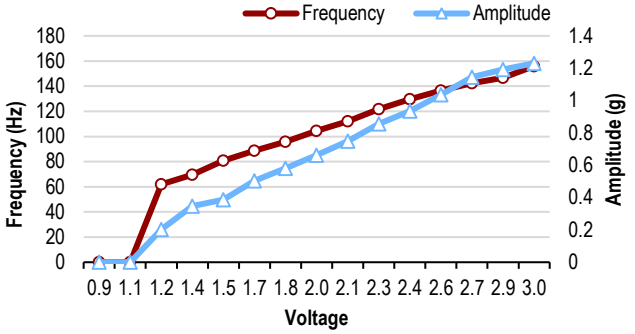


Figure 3. Measured frequency and amplitude of an ERM vibro-motor mounted on a user's wrist as applied voltage changes. Graph adapted from our prior work [18].

To provide corrective guidance, no vibration occurs if the difference between the finger's trajectory and the intended direction is less than 22.5°; 22.5° was selected based on prior work showing the minimum angular accuracy in moving the hand based on vibration around the wrist is at least 23° [18]. As the finger's trajectory diverges from the intended direction, vibration intensity increases: the applied voltage increases linearly from 1V when the user's hand movement trajectory is 22.5° away from the intended direction to 3V (maximum intensity) when the user's trajectory is 90° or farther from the intended direction.

3.2.2 Interpolated Vibration

With the second feedback design, one or two motors vibrate at a time to allow the user to interpolate a more precise direction. Vibrating two adjacent motors should theoretically create an illusory single 'phantom sensation' between the two [2]; varying the amplitude and frequency of each motor impacts the specific location of the phantom sensation. Following [2], we logarithmically adjust the amplitude and frequency based on the precise angle we wish to indicate between the motors. To indicate an angle θ between two motors, we calculate:

$$V_1 = V_{max} \times \frac{log(\theta_2 - \theta + 1)}{log(\theta_2 - \theta_1 + 1)}$$

$$V_2 = V_{max} \times \frac{log(\theta - \theta_1 + 1)}{log(\theta_2 - \theta_1 + 1)}$$

where V_1 and V_2 are the voltages on the two adjacent motors and θ_1 and θ_2 are the angles of directions mapped to the two motors as shown in the Figure 2. As with the single motor vibration design, V_{max} was controlled to avoid continuous vibration on the wrist. V_{max} was set to 3V if the difference between the intended direction and the finger movement was greater than 90°. If the difference was less than 22.5°, V_{max} was 0V. If the difference was in the 22.5–90° range, V_{max} linearly ranged from 1V–3V.

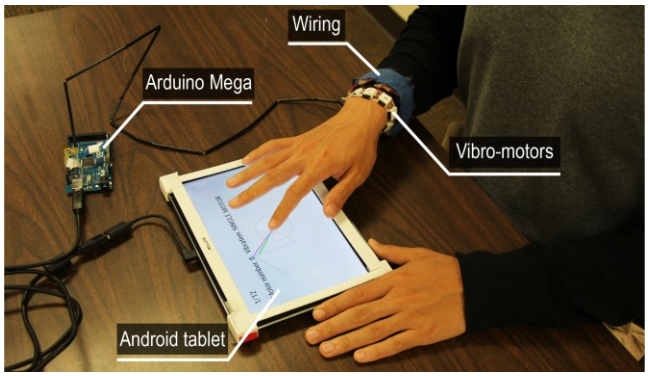


Figure 4. The experimental setup for both studies. Visuals on the screen were for the experimenter and were not visible to sighted participants nor perceptible to blind participants.

4. STUDY 1: INTERPOLATION AND NUMBER OF MOTORS

To gather preliminary data before running a full study with blind and visually impaired participants, we first conducted an initial study with 11 sighted and 2 blind participants. This study included both the 4-motor and 8-motor wristbands as well as the single-motor and interpolated feedback. It was not meant to substitute for evaluation with blind and visually impaired participants but instead allowed us to identify the most promising feedback designs to evaluate later with our VI participants—thus allowing for a simplified and more statistically powerful experimental design.

4.1 Method

4.1.1 Participants

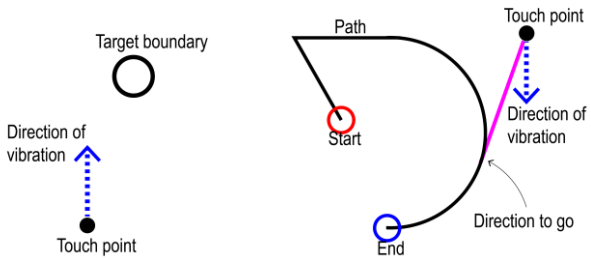
We recruited 11 sighted participants (6 female, 5 male) and two blind participants.[2] The sighted participants were recruited through campus email lists and ranged in age from 19 to 34 (M = 25.3, SD = 4.5); all were right-handed. The blind participants were recruited through a list of potential participants maintained by our research lab. One was 53 years old, male, and right-handed, while the second was 63 years old, female, and left-handed.

4.1.2 Procedure

The study session took up to two hours. We first collected demographic information and touchscreen device experience. The tablet device, which had non-slip feet attached, was placed flat on a table in front the participant, as shown in the Figure 4. The four conditions were presented in counterbalanced order with a five-minute break offered between each one. Sighted participants were blindfolded during the tasks to isolate the effects of the haptic feedback and to provide a better indication of which conditions we should evaluate in the follow-up study with VI participants. The procedure for each condition consisted of the following four steps.

Step 1: Wristband placement and introduction. The experimenter placed the wristband on the participant's wrist and adjusted the position of the motors while the participant held their hand on the tablet screen. To confirm that the motors were perceptually at the correct positions (Figure 2), the experimenter vibrated each one in turn and adjusted its position if the participant reported that it felt off. Participants were then given a brief introduction to the wristband (~5 minutes) that consisted of 16 trials: a motor vibrated, the participant attempted to move their finger in

[2] An additional five sighted participants completed the study but their data was not usable due to a technical problem with Bluetooth where the Arduino and tablet disconnected repeatedly. In Study 2, we resolved this problem by using a wired connection.



(a) Target-finding task **(b) Path-tracing task**

Figure 5. Single-motor haptic feedback for example target-finding and path-tracing trials. The motor closest to the ideal angle of movement is vibrating. For the path-tracing task, the ideal angle of movement is computed using a tangent line drawn from the touch point to the curve.

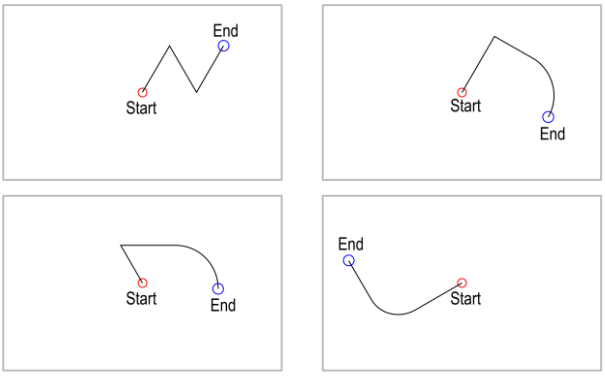


Figure 6. Sample paths used in the path-tracing task. All paths started at the center of the screen.

the direction of the vibration, and the system provided feedback about how close the movement was to the presented angle of movement (perfect: ≤ 5.625° from presented angle, close: > 5.625° and ≤ 22.5°, and incorrect: > 22.5°). If the movement was incorrect, the direction was repeated until the participant moved correctly, with the experimenter giving verbal instructions after two incorrect attempts. For the single-motor feedback conditions, the 16 trials consisted of 4 or 2 trials per motor for the 4-motor and 8-motor wristbands, respectively. For the interpolated feedback conditions, 16 target directions were evenly distributed at 22.5° intervals for both wristbands, with one trial in each direction.

Step 2: Target-finding task (Figure 5a). For the target-finding task, each trial consisted of moving from the center of the touchscreen to a randomly placed circular target (15mm in diameter) with at least 7.5mm buffer (a target radius distance) between the edge of the target and the edge of the screen. A sticker at the center of the tablet provided a tactile indicator of the start location. To begin a trial, the participant placed their finger on the sticker for one second, after which a sound played and wristband vibration indicated the trial had begun. Upon entering the target bounds, an end sound also played. Participants completed 5 training trials and 30 test trials for this task, where the training trials included decreasing levels of experimenter guidance, from physically guiding the participant's finger to having the participant do the trial fully independently. Participants were asked to complete the test trials as quickly and accurately as possible.

Step 3: Path-tracing task (Figures 5b and 6). Following the target-finding task, participants were asked to trace paths (as if they were following a route on a map). We programmatically generated a set of 18 paths based on route features from a real university campus map, including the relative length of segments, the angles between segments, and the ratio of straight to curved segments. Each path consisted of three straight or curved segments, with the constraint that not all three could be curved. The distance between the start and end points of each path was at least 30 mm. The average path length was 140 mm (*SD*=26.4).

During this portion of the study, each participant completed 5 path-tracing training trials presented in the same order followed by 10 test trial paths presented in random order. As with target finding, the experimenter provided decreasing levels of guidance during the training trials. Again, a trial started by placing the finger on the start sticker until a sound played and vibration started. The system then guided the user to trace each segment in turn. For straight segments, the system guided the user to the segment end point in the same manner as guiding to a target in the target-finding task. For curved segments, the system guided the user to the tangent point of a line drawn from the curve to the finger, the location of which was continuously updated as the finger moved (Figure 6). For both straight and curved segments, reaching the "end of the segment" was defined as being within 7.5 mm of the exact end point (*i.e.,* within the radius of a target in the target-finding task). A beep indicated the end of each segment and a chime sound indicated the end of the trial. Again, participants were asked to complete the test trials as quickly and accurately as possible.

Step 4: Subjective feedback. After each condition (*e.g.,* single *vs.* interpolated with 4 *vs.* 8 motors), we asked participants to rate overall experience, ease, accuracy, and speed on 7-point Likert scales, and solicited open-ended feedback. When all conditions were done, we asked participants to select the easiest, fastest, most accurate, most preferred, and least preferred conditions and to provide overall feedback about their experience.

4.1.3 Study Design

The study used a within-subjects design with two factors: *Number of Motors* (4-motor *vs.* 8-motor) and *Vibration Type* (single motor *vs.* interpolation). The order of presentation for the four conditions was counterbalanced using a balanced Latin square and participants were randomly assigned to orders.

4.1.4 Measures and Data Analysis

The main measures were trial completion time and movement error. Trial completion time was computed from the time the vibration started until their finger entered the target bounds for target-finding trials or reached the end of the last segment for path-tracing trials. For target-finding trials, movement error was computed as the ratio of: the actual distance the finger moved during the trial to the Euclidian distance between the start location and the closest point on the target (lower values are better):

$$movement\ error = \frac{finger\ movement\ distance}{shortest\ possible\ distance}$$

For path-tracing trials, movement error was computed by first uniformly resampling the user's actual trace to have one point every 1 mm, then matching the points from the resampled path to the presented path using dynamic time warping (DTW) [6], which measures the distance between two temporal sequences of points. The average of the distances between matched points was then computed as the movement error; lower numbers are better.

For each participant, outlier trials were defined based on trial completion time as being more than 3*IQR (interquartile range) above the upper quartile or less than 3*IQR below the lower quartile within a given experimental condition [39]. In total, 1.6% of trials were excluded from analysis. The trial completion times and movement errors for both tasks violated the normality assumption (Shapiro-Wilk tests, $p < .05$). Therefore, we ran nonparametric analyses using 2-way ANOVA with aligned rank transform (ART) [42]. Bonferroni corrections were used for all posthoc pairwise comparisons.

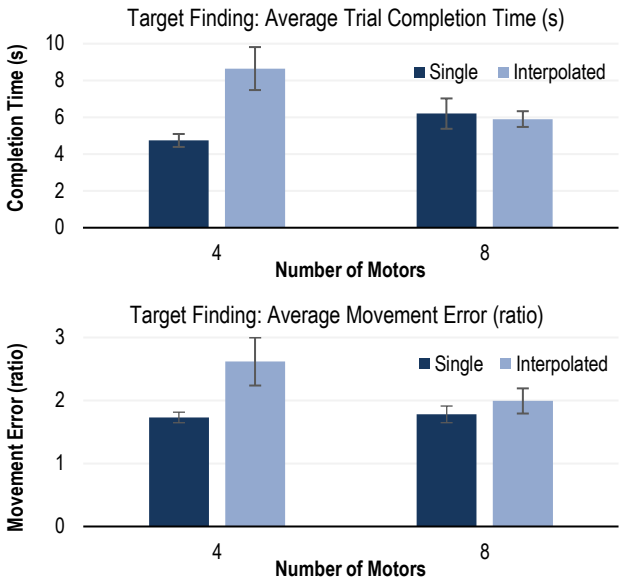


Figure 7. Average trial completion time and movement error for the target-finding task in Study 1, showing interpolated feedback is less accurate than single-motor feedback. Error bars show standard standard error. (*N* = 13)

4.2 Results

We present speed and accuracy for both the target finding and path-tracing tasks, followed by subjective feedback.

4.2.1 Target Finding

Figure 7 (top) shows the average trial completion times for the target-finding task. Overall, the interpolated feedback was slower than the single-motor feedback, at 7.3s per trial on average (*SD* = 3.4) compared to 5.4s (*SD* = 2.3). This difference was significant, as shown by a main effect of *Vibration Type* ($F_{1,36} = 8.83$, $p = .005$, $\eta^2 = 0.21$). While there was no significant effect of *Number of Motors* ($F_{1,36} = 0.53$, $p = .470$, $\eta^2 = 0.04$), there was a significant interaction effect between *Number of Motors* and *Vibration Type*, showing that efficacy of the vibration type depended on how many motors there were ($F_{1,36} = 8.91$, $p = .005$, $\eta^2 = 0.13$). Pairwise comparisons showed that interpolated feedback was slower than single-motor feedback with the 4-motor wristband, at 8.6s per trial on average (*SD* = 4.2) compared to only 4.7s per trial (*SD* = 1.3); this difference was significant with a Wilcoxon signed rank test ($W = 3$, $Z = -2.97$, $p = .001$, $r = 0.58$). In contrast, no difference was found between the two vibration types for the 8-motor wristband with 6.2s per trial (*SD*=3.1) for single motor and 5.9s (*SD*=1.5) for interpolated ($W = 42$, $Z = -0.24$, $p = .839$, $r = 0.05$).

In terms of accuracy (Figure 7, bottom), interpolated feedback also resulted in higher error than single-motor feedback, with a movement error of 2.3 on average (*SD* = 1.1) compared to 1.8 (*SD* = 0.4); this difference was significant, as shown by a main effect of *Vibration Type* ($F_{1,36} = 5.67$, $p = .023$, $\eta^2 = 0.14$). The main effect of *Number of Motors* ($F_{1,36} = 1.71$, $p = .200$, $\eta^2 < 0.01$) and the

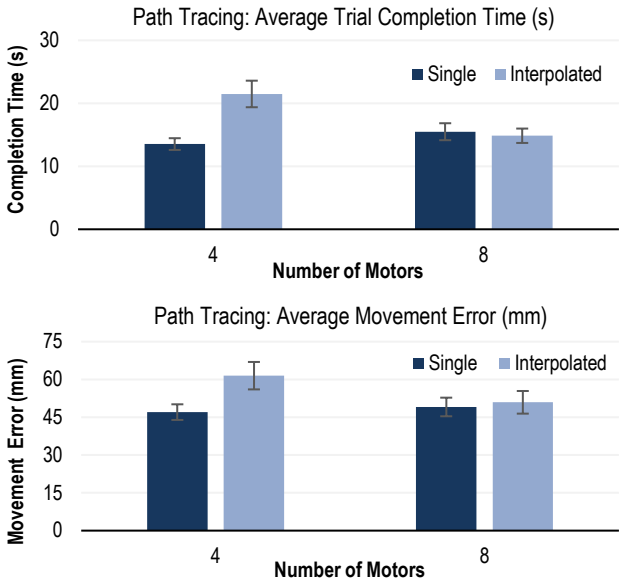


Figure 8. Average trial completion time and movement error for the path-tracing task in Study 1, showing interpolated feedback is slower and less accurate than single-motor feedback. The error bars are standard error. (N = 13)

interaction between *Number of Motors* and *Vibration Type* ($F_{1, 36}$ = 1.62, p = .212, η^2 = 0.04) were not significant.

4.2.2 Path Tracing

This pattern of findings was similar for the path-tracing task, with interpolated feedback not faring well, particularly with the 4-motor wristband design. Figure 8 (top) shows the average trial completion times. Interpolated feedback was significantly slower than single-motor feedback with average trial times of 18.1s (SD = 6.9) compared to 14.5s (SD = 4.2); main effect of *Vibration Type* ($F_{1, 36}$ = 8.27, p = .007, η^2 = 0.24). The number of motors did not have a significant effect on the trial completion time ($F_{1, 36}$ = 4.93, p = .057, η^2 = 0.14); however, there was a significant interaction effect between the *Number of Motors* and the *Vibration Type* ($F_{1, 36}$ = 12.5, p = .001, η^2 = 0.15). As with the target-finding task, pairwise comparisons showed that this latter result was due to interpolated feedback slowing participants down only with the 4-motor wristband. Here, trials took on average 21.5s (SD = 7.6) with interpolated feedback and 13.5s (SD = 3.4) with single-motor feedback, which was a significant difference (W = 90, Z = 3.11, p < .001, r = 0.61). There was no such difference for the 8-motor wristband.

Movement error (Figure 8, bottom) was also significantly higher with interpolated feedback than with single-motor feedback, at 9.5mm on average (SD = 3.1) compared to 8.2mm (SD = 2.1); main effect of *Vibration Type* ($F_{1, 36}$ = 4.66, p = .038, η^2 = 0.11). While the main effect of *Number of Motors* was not significant ($F_{1, 36}$ = 2.2, p = .147, η^2 = 0.11), the interaction effect between *Number of Motors* and *Vibration Type* was significant ($F_{1, 36}$ = 4.17, p = .048, η^2 = 0.04). As with earlier results, interpolated feedback resulted in higher movement error than single-motor feedback for the 4-motor wristband, at 10.6mm (SD = 3.3) versus 8mm (SD = 1.9); the difference was significant (W =80, Z = 2.41, p = .013, r = 0.47).

4.2.3 Subjective Responses

Subjective responses reflected the performance results, as shown in Figure 9. Participants consistently rated the 4-motor wristband with interpolation worse than the other three conditions in terms of ease, speed, and accuracy. Table 1 shows vote tallies for easiest, fastest, most accurate, and most/least preferred of the four conditions. The single-motor feedback options received more positive votes compared to interpolated feedback options.

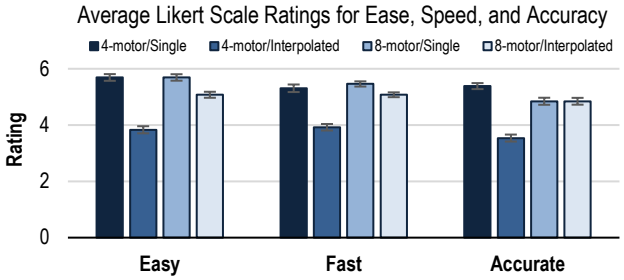


Figure 9. Study 1 subjective ratings on 7-point Likert scales for ease, speed, and accuracy, where 1=*strongly disagree* and 7=*strongly agree*. The single motor condition received higher ratings than interpolated feedback. Error bars are standard error. (N = 13)

	4-motors		8-motors	
	Single	Interpolated	Single	Interpolated
Easiest	7	1	4	1
Fastest	6	1	4	2
Most accurate	9	1	3	0
Most preferred	6	2	4	1
Least preferred	0	9	1	3

Table 1. Subjective vote tallies in Study 1, showing 4-motors with interpolation was the least preferred condition. (N = 13)

4.3 Summary and Discussion

Interpolated feedback was slower and less accurate in both tasks, but was particularly problematic with the 4-motor wristband. Subjective feedback matched this performance data. One limitation of this study is that the sighted participants' performance may not be representative of blind users. However, performance from the two blind participants matched the overall trends comparing single-motor *vs.* interpolated feedback: both blind participants were faster with single-motor feedback for both tasks, while single-motor feedback was more accurate for path tracing for both participants and resulted in mixed accuracy for target finding (one participant was more accurate, one was less accurate). Based on these results, we focused only on single-motor feedback in Study 2 with blind and VI participants.

5. STUDY 2: EVALUATION WITH BLIND AND VISUALLY IMPAIRED USERS

Based on the findings from Study 1, we eliminated the interpolated feedback conditions and conducted a follow-up study with 14 blind and visually impaired participants to compare single-motor feedback with the 4-motor and 8-motor wristbands.

5.1 Method

The study method is largely the same as for the pilot study with a few key changes highlighted below.

5.1.1 Participants

We recruited 14 (8 female, 6 male) visually impaired participants through an existing participant pool and local organizations working with people with visual impairments. Participants were on average 44.8 years old (SD=13.9; range 22–64). Seven participants were totally blind, two were blind with light perception, and five were legally blind. All 14 used a screen reader on their computer and/or smartphone. Twelve participants were right handed, one was left handed, and one reported using her left hand for writing tasks

and right hand for touchscreen devices (she used her right hand for study tasks). All but one participant owned a touchscreen device. Of the 13 touchscreen device owners, 11 participants reported daily use and 2 participants reported use every few days.

5.1.2 Procedure

The study procedure lasted up to 90 minutes. As we removed the interpolated feedback conditions, we had only two experimental conditions: single-motor feedback with four or eight motors. The conditions were fully counterbalanced and participants completed the same set of tasks with each condition as in Study 1. Due to a steep learning curve that we saw persist into the test trials during Study 1, we also provided more training and increased the number of test trials as follows: (1) After first putting on the wristband, participants completed the same 16 *familiarization* trials as in Study 1; however, an additional trial was appended every time a participant made two consecutive incorrect attempts in a specific direction, up to a maximum of four or eight additional trials depending on the number of motors on the wristband. (2) For the *target-finding* task, participants completed at least 12 training trials including an initial two guided by the experimenter. Additional training trials were added up to a maximum of 20 if the movement error (ratio) was greater than 3.0 in two or more of the most recent five training trials. The number of test trials for target finding also increased from 30 in Study 1 to 36 here. (3) For the *path-tracing* task, participants completed 7 training trials and 12 test trials (up from 5 and 10, respectively in Study 1). (4) Finally, because of comments from some Study 1 participants about desensitization to the haptic vibration, we imposed an eight-minute break between conditions.

5.1.3 Study Design, Measures and Data Analysis

Study 2 included a single experimental factor of *Number of Motors* (4 or 8). This was a within-subjects factor and was fully counterbalanced, with participants randomly assigned to orders. However, to capture possible learning effects, we split the 36 trials in the target-finding task into two blocks of 18 trials each and included *Block* as an additional factor in the analysis of that task. We did not do so for the path-tracing task because of the smaller number of trials and the increased likelihood of any individual trial unduly affecting the average within a block of only nine trials.

As with Study 1, the main measures were *trial completion time* and *movement error*. Outlier trials were detected within each condition per participant using the IQR method as in Study 1. Out of 1,344 test trials in total across the 14 participants, 31 outlier trials (2%) were excluded from data analysis.

Shapiro-Wilk tests showed that trial completion times and movement error violated the normality assumption of parametric tests for both tasks ($p < .05$). Thus, for the target-finding task, a 2-way repeated measures ANOVA (*Number of Motors* x *Block*) with ART was used to anlayze the trial completion time and movement error. We used Wilcoxon signed rank tests to analyze the trial completion time and movement error for the path-tracing task, as well as for subjective ratings.

5.2 Results

We present performance analyses followed by subjective ratings and an analysis of open-ended comments from participants.

5.2.1 Target Finding

As shown in Figure 10 (top), similar to Study 1, participants were again faster with the 4-motor wristband than the 8-motor one, at 7.3s per trial on average ($SD = 4.8$) compared to 9.8ms ($SD = 4.9$). This difference was significant (main effect of *Number of Motors*: $F_{1, 39} = 14.51$, $p < .001$, $\eta^2 = 0.05$). The main effect of *Block* was

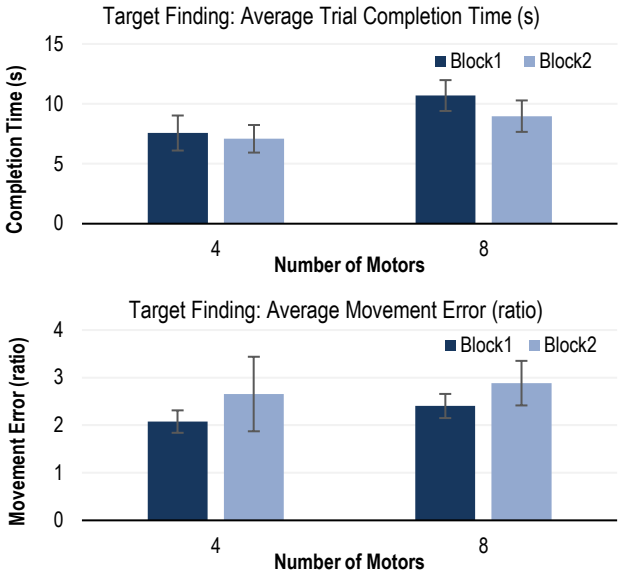


Figure 10. Trial completion time and movement error for the target-finding task in Study 2. The 4-motor wristband was faster and more accurate than the 8-motor wristband. Error bars are standard error. ($N = 14$)

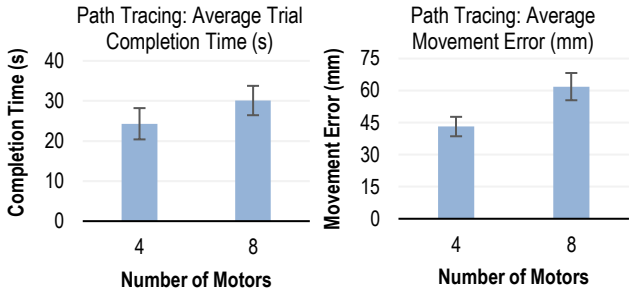


Figure 11. Trial completion time and movement error for the path-tracing task in Study 2. The 4-motor wristband was more accurate than the 8-motor wristband. Error bars are standard error. ($N = 14$)

not statistically significant ($F_{1,39} = 2.54$, $p = .119$, $\eta^2 < 0.01$) nor was the interaction between the two factors ($F_{1, 39} = 1.45$, $p = .235$, $\eta^2 < 0.01$).

Participants were also more accurate with the 4-motor wristband than the 8-motor one (Figure 10, bottom). The movement error ratio was on average 2.4 with four motors ($SD = 2.1$) compared to 2.6 with eight motors ($SD = 1.4$), a significant difference (main effect of *Number of Motors*: $F_{1, 39} = 5.14$, $p = .029$, $\eta^2 < 0.01$). As with completion time, the main effect of *Block* was not statistically significant ($F_{1, 39} = 0.34$, $p = .560$, $\eta^2 = 0.01$) nor was the interaction between the two factors ($F_{1,39} < 0.01$, $p = .947$, $\eta^2 < 0.01$).

5.2.2 Path Tracing

Performance for the path-tracing task is shown in Figure 11. The results partially mirror the target-finding results: participants were not faster with either wristband ($W = 25$, $Z = 1.73$, $p = .091$, $r = 0.33$) but were significantly more accurate with the 4-motor wristband ($W = 12$, $Z = -2.54$, $p = .009$, $r = 0.48$). Trials took on average 24.3s with four motors ($SD = 14.6$) and 30.1s with eight motors ($SD = 13.7$). The movement error was 7.4 mm ($SD = 2.9$) with four motors and 10.6 mm ($SD = 4.1$) with eight motors on average.

	4-motor *M(SD)*	8-motor *M*(*SD*)	Wilcoxon signed rank test result
Ease of understanding*	6.4 (0.7)	5.2 (1.6)	W=36, z = -2.54, p = .008, r=0.48
Speed	6.1 (0.8)	5.1 (1.2)	W = 61, z = -1.75, p = .085, r = 0.33
Accuracy*	5.9 (1.0)	4.8 (1.1)	W = 59, z = -2.36, p = .021, r = 0.45

Table 2. Subjective ratings on 7-point Likert scales, were 7 indicated strong positive agreement. Measures with statistically significant differences are shown with '*'.

5.2.3 Subjective Feedback

Table 2 shows subjective Likert scale ratings collected at the end of each condition, which largely reflect the performance results. Participants perceived the 4-motor wristband to be significantly easier to understand and more accurate than the 8-motor wristband. No differences were found in perceptions of speed.

When asked about overall preference at the end of the study, participants were split: eight participants preferred the 4-motor feedback while six participants preferred the 8-motor feedback. The most common reason cited for preferring 4-motor feedback was that it was easier to understand (6 participants). For example, P9 said: "*[4-motors] was easier to use and I felt less frustrated. I felt like I did better. I was more sure of […] which one was vibrating."* For the 4-motor wristband, some participants mentioned the ease of mapping the vibrations to the four cardinal directions on the tablet screen (one person mentioned this for 8-motor feedback too). One participant also mentioned that with four motors it felt like there was less vibration, whereas with eight motors it felt like the entire hand was vibrating. Finally, one participant appreciated the spacing between motors with the 4-motor option because she had a small wrist.

Reasons for preferring 8-motor feedback included higher perceived accuracy and increased precision. For example, P8 said, "*The feedback is more fine-grained and I like that [...] Instead of a general direction I like precision."* P3 also mentioned:

> *"...this would be good for [...] not just maps but statistics, because [...] when you put the numbers on a graph, you know, you have more, more to figure out where everything is for charting the graph."* (P3)

At the same time P11 noted the additional information with the 8-motor feedback had some drawbacks, saying: "*I could adjust better [with 8-motor], but I did think that it slowed me down more."*

Some participants mentioned applications for which they thought the wrist-worn haptic feedback could be useful, including: drawing, navigating maps, plotting graphs, and operating touchscreen devices. At the same time, issues of learning and fatigue arose, suggesting that the haptic designs could be further refined to limit vibration. Eight participants commented that they would have performed better with more practice. For example. P15 mentioned for 8-motor feedback: *"I feel it, I understand. [The] more I was practicing, [the] more I was getting it."* Four participants mentioned that after a while their wrist got tired, and as a result, they had more trouble identifying which motor was vibrating. For example, P07 mentioned for 8-motor feedback: *"After a while [...] the way it jumps [from one direction to another], if there is no constant stopping of vibrations, it kind of starts blending into each other."* We return to this issue in the Discussion section.

5.3 Summary

Study 2 largely reaffirms our Study 1 findings. Participants completed the target-finding task faster and more accurately with the 4-motor wristband than the 8-motor wristband. In the path-tracing task, participants were significantly more accurate when they used 4-motor wristband than 8-motor wristband. The subjective evaluation supported the performance results with positive feedback about the 4-motor wristband.

6. DISCUSSION

Across the two experiments, our findings showed that the fastest, most accurate, and most preferred design was the four-motor wristband with a single-motor vibration. This conclusion is counter to our expectations: we had expected that having more vibromotors (8 *vs.* 4) and attempting to induce precise phantom sensations between adjacent motors would lead to more accurate performance. We discuss this and other issues relevant to future wrist-worn haptic feedback designs below.

6.1 Why Does 4-motor Outperform 8-motor?

We were surprised that doubling the number of haptic actuators around the wrist did not result in performance benefits: the 8-motor design was slower and less accurate than the 4-motor design in both studies. While further work is necessary to uncover why this might be, some possible reasons include vibration transfer, higher cognitive load, and perceptual limitations. Vibration transfer occurs when the actuation of one motor causes one or more adjacent motors to also vibrate. The transfer is a function of the intensity of the actuated motor, the distance and physical connection between the motors, and the wrist itself (*e.g.,* the skin, bone structure). While our elastic band was designed to mitigate vibration transfer, it was not completely suppressed. In particular, because the inter-motor distance was lower with the 8-motor design, transfer would theoretically be higher. The interpolated conditions also used higher combined intensity vibrations than the single-motor conditions, which may have exacerbated transfer problems and created inadvertent phantom sensations or other perceptual issues. In practice, embedding the motors in a solid smartwatch strap or wristband robust enough for daily use will likely result in greater vibration transfer than our elastic design, magnifying the benefits of the 4-motor design over the 8-motor design.

6.2 Designing Wristband Haptics

Creating wrist-worn directional hand guidance systems requires exploring a large design space. In this paper, we built and investigated only a few possible designs. Important design considerations, include: cognitive factors (*e.g.,* understandability, learnability), perceptual factors (*e.g.,* desensitization, phantom sensations), as well as the design of the wristband, haptic hardware, and underlying software algorithms to actuate those haptics.

Early in our design process, we examined a range of haptic hardware, including muscle wire (shape memory Nitinol), tactors, piezoelectrics, and vibromotors. We settled vibromotors based on our internal experiments and results from prior work (*e.g.,* [18]). In the haptics literature, two vibromotor designs are common: ERM and LRA. We initially built and examined prototypes for both. While LRAs are attractive because they offer faster response times than ERMs and allow for independent control of amplitude and frequency, our internal tests showed that our ERM design performed better. For example, in a small pilot test with three participants, all participants performed faster and more accurately with the ERM prototype than the LRA prototype. Consequently, we refined and used the ERM prototype in this paper.

As with the haptic hardware, designing an appropriate haptic feedback signal is a difficult design problem. One needs to consider the hardware response time, the intensity and frequency of the

vibration, and the vibro-modulation pattern. To help inform the approach used in this paper, we designed four guidance methods. Three of the four methods used *corrective guidance*—that is, the vibromotors were actuated only when the user's finger trajectory and the intended direction differed by at least 22.5°. Of these corrective guidance approaches, one vibrated at maximum intensity regardless of the magnitude of the difference, one increased linearly as the angular difference increased, and the third increased in a small number of discrete steps as the angular difference increased. A fourth option was to use a pulse vibration (100ms on / 100ms off) to indicate intended direction, regardless of the difference between the actual and intended direction. Among the four methods, the corrective guidance with linear change in vibration was fastest and most accurate in our pilot tests, so it was selected for our studies. However, further work may yield improved haptic feedback designs (*e.g.,* more perceptible feedback patterns).

6.3 Effect of Age and Technology Experience

Before refining our prototypes and approach for Study 2, we attempted to replicate Study 1 with 22 blind participants (M=54.3 years old; SD=10). In contrast to our Study 1 and Study 2 experiences, only 10 of the 22 participants in this interim study could complete the experiment within the allotted time; the other 12 dropped out early or were unable to complete all tasks. This experience partly motivated us to increase the number of training and test trials in Study 2. At the same time, participant age and overall familiarity with technology may have played a role in the low success rates in this interim study. The interim study participants were older than in Studies 1 and 2: on average 54.3 years old (compared to 28.3 in Study 1 and 44.8 in Study 2). Age is an important variable given that older persons have lower sensitivity for perceiving vibration stimulus on the skin [11]. The interim participants also had less experience with touchscreen technology, with only 14 of the 22 owning a touchscreen device, compared to almost all participants in Studies 1 and 2.

6.4 Limitations and Future Work

The results and limitations of our studies reveal several directions for future work. First, some participants mentioned fatigue from using the haptic feedback, which may be partly due to desensitization. While we attempted to mitigate desensitization by providing corrective guidance rather than continuous vibration (Section 3.2), it could have impacted some feedback designs more than others (*e.g.*, interpolated *vs.* simple); for example, desensitization occurs more quickly at higher intensity levels [5] and with more sustained activation [23]. We also included 5-8 minute breaks between conditions, but recovery time between vibration stimuli is still an active area of research (most work suggests 30 secs to 3 mins, *e.g.*, [5, 14, 23]). Ultimately, to limit fatigue, wrist-based haptic feedback may be most applicable to short tasks (*e.g.*, learning a new route, exploring a drawing, reading a flyer) rather than in-depth tasks (*e.g.*, reading a book line-by-line). It is also possible that fatigue lessens as participants gain experience. Future work should explore these concerns further.

Second, some participants indicated needing more time to learn and effectively use our wristband prototypes. While we examined learnability by splitting Study 2 into blocks, we found no clear learning effect during the duration of the target-finding task. Future work should examine how performance changes with time by using a more longitudinal study design.

Third, while our study tasks were developed to model two common real-world tasks, finding a target and tracing a route, more work is necessary to understand whether the observed performance numbers are accurate enough to provide value in practice (*e.g.*, did participants have a strong conceptual understanding of the paths that they tried?). It will also be important to explore other applications, such as drawing, plotting graphs, and operating touchscreen devices in general, as suggested by participants.

Fourth, our focus was on isolating the effects of specific wrist-based haptic designs but non-haptic and hybrid approaches should also be considered (*e.g.*, sonification or vocal directions like "move up"). Note, however, that the audio channel is often overloaded in assistive applications and audio directional guidance can interfere with speech output (*e.g.*, screenreaders) [38].

7. CONCLUSION

In this paper, we report on the design and evaluation of two haptic wristbands, 4-motor *vs.* 8-motor, and two actuation approaches, simple (single motor) *vs.* interpolated (two motor). Our Study 1 findings showed that participants could non-visually find targets and trace paths more quickly and accurately with single-motor feedback than with interpolated feedback, particularly with the 4-motor design. In Study 2, we eliminated the interpolated feedback conditions and found that single-motor feedback with four motors was faster, more accurate, and most preferred by blind participants compared to similar feedback with eight motors. Our results help establish important benchmarks for future work and contribute to the growing body of hand guidance research.

8. ACKNOWLEDGMENTS

We thank Hong Tan, David Ross, and the Atlanta VAMC. This work was funded by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

9. REFERENCES

[1] Aggravi, M., Salvietti, G. and Prattichizzo, D. 2016. Haptic wrist guidance using vibrations for Human-Robot teams. *IEEE 25th Intl Symposium on Robot and Human Interactive Communication (RO-MAN)*, 113–118.

[2] Alles, D.S. 1970. Information Transmission by Phantom Sensations. *IEEE Transactions on Man-Machine Systems*. 11, 1 (1970), 85–91.

[3] Amemiya, T., Yamashita, J., Hirota, K. and Hirose, M. 2004. Virtual leading blocks for the deaf-blind: A real-time way-finder by verbal-nonverbal hybrid interface and high-density RFID tag space. *Virtual Reality Annual International Symposium*. (2004), 165–172.

[4] Arter, C., McCall, S. and Bowyer, T. 1996. Handwriting and children with visual impairments. *British Journal of Special Education*. 23, 1 (1996), 25–28.

[5] Berglund, U. and Berglund, B. 1970. Adaptation and recovery in vibrotactile perception. *Perceptual and motor skills*. 30, 3 (1970), 843–853.

[6] Berndt, D. & Clifford, J. 1994. Using dynamic time warping to find patterns in time series. *KDD workshop'94*, 359–370.

[7] Bosman, S., Groenendaal, B., Findlater, J., Visser, T., de Graaf, M. and Markopoulos, P. 2003. GentleGuide: An exploration of haptic output for indoors pedestrian guidance. *HCI with Mobile Devices and Services'03*, 358–362.

[8] Brock, A., Kammoun, S., Macé, M. and Jouffrais, C. 2014. Using wrist vibrations to guide hand movement and whole body navigation. *i-com*. 13, 3 (Jan. 2014), 19–28.

[9] Carcedo, M.G., Chua, S.H., Perrault, S., Wozniak, P., Joshi, R., Obaid, M., Fjeld, M. and Zhao, S. 2016. Hapticolor:

Interpolating color information as haptic feedback to assist the colorblind. *Proc. of CHI'16*, 3572–3583.

[10] Chen, H.Y., Santos, J., Graves, M., Kim, K. and Tan, H.Z. 2008. Tactor localization at the wrist. *Lecture Notes in Computer Science*. 5024 LNCS, (2008), 209–218.

[11] Cholewiak, R.W. and Collins, A. a 2003. Vibrotactile localization on the arm: effects of place, space, and age. *Perception {&} psychophysics*. 65, 7 (2003), 1058–1077.

[12] Cosgun, A., Sisbot, E.A. and Christensen, H.I. 2014. Evaluation of rotational and directional vibration patterns on a tactile belt for guiding visually impaired people. *IEEE Haptics Symposium, HAPTICS*. (2014), 367–370.

[13] Crossan, A. and Brewster, S. 2008. Multimodal trajectory playback for teaching shape information and trajectories to visually impaired computer users. *ACM Trans. Access. Comput.* 1, 2 (Oct. 2008), 12:1--12:34.

[14] van Erp, J.B.F. 2002. Guidelines for the Use of Vibro-Tactile Displays in Human Computer Interaction. *Proceedings of Eurohaptics*. (2002), 18–22.

[15] Ertan, S., Lee, C., Willets, A., Tan, H. and Pentland, A. 1998. A wearable haptic navigation guidance system. *Proc. of ISWC'98*. 164–165.

[16] Goldreich, D. & Kanics, I. 2003. Tactile Acuity is Enhanced in Blindness. *J. Neurosci.* 23, 8 (Apr. 2003), 3439–3445.

[17] Hirose, M. and Amemiya, T. 2003. Wearable Finger-Braille interface for navigation of Deaf-Blind in ubiquitous barrier-free space. *Proc. HCI International'03*, 1417–1421.

[18] Hong, J., Stearns, L., Froehlich, J., Ross, D. and Findlater, L. 2016. Evaluating Angular Accuracy of Wrist-based Haptic Directional Guidance for Hand Movement. *Proceedings of the 42nd Graphics Interface Conference* (2016), 195–200.

[19] How to Make a Sliding Knot (single knot) - jewelry making tutorial: 2012. *https://youtu.be/t6Qy5WdSq74*.

[20] Kammoun, S., Jouffrais, C., Guerreiro, T., Nicolau, H. and Jorge, J. 2012. Guiding Blind People with Haptic Feedback. *Pervasive 2012 Workshop on Frontiers in Accessibility for Pervasive Computing* (Jan. 2012).

[21] Kim, K., Ren, X., Choi, S. and Tan, H.Z. 2016. Assisting people with visual impairments in aiming at a target on a large wall-mounted display. *International Journal of Human-Computer Studies*. 86, (2016), 109–120.

[22] Lee, J., Han, J. and Lee, G. 2015. Investigating the Information Transfer Efficiency of a 3x3 Watch-back Tactile Display. *Proc of CHI'15*, 1229–1232.

[23] Leung, Y.Y., Bensmaia, S.J., Hsiao, S.S. and Johnson, K.O. 2005. Time-course of vibratory adaptation and recovery in cutaneous mechanoreceptive afferents. *Journal of neurophysiology*. 94, 5 (2005), 3037–3045.

[24] Mann, S., Huang, J., Janzen, R., Lo, R., Rampersad, V., Chen, A. and Doha, T. 2011. Blind navigation with a wearable range camera and vibrotactile helmet. *Proc. of MM'11*, 1325-1328.

[25] Matscheko, M., Ferscha, A., *et al.* 2010. Tactor placement in wrist worn wearables. *Proc. of ISWC'10*, 1–8.

[26] Oh, U., Kane, S.K. and Findlater, L. 2013. Follow That Sound: Using Sonification and Corrective Verbal Feedback to Teach Touchscreen Gestures. *Proc. of ASSETS'13*, 13:1--13:8.

[27] Panëels, S., Brunet, L. and Strachan, S. 2013. Strike a Pose: Directional Cueing on the Wrist and the Effect of Orientation. *Haptic and Audio Interaction Design*. I. Oakley and S. Brewster, eds. Springer Berlin Heidelberg. 117–126.

[28] Parseihian, G., Gondre, C., Aramaki, M., Ystad, S. and Kronland-Martinet, R. 2016. Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Transactions on Multimedia*. 18, 4 (2016), 674–686.

[29] Pasquero, J., Stobbe, S. & Stonehouse, N. 2011. A haptic wristwatch for eyes-free interactions. *CHI'11*, 3257–3266.

[30] Richter, H., Blaha, B. and Wiethoff, A. 2011. Tactile feedback without a big fuss: simple actuators for high-resolution phantom sensations. *Proc. of UbiComp '11*. (2011), 85–88.

[31] Sampaio, E., Maris, S. and Bach-y-Rita, P. 2001. Brain plasticity: "visual" acuity of blind persons via the tongue. *Brain research*. 908, 2 (Jul. 2001), 204–7.

[32] Scheggi, S., Talarico, A. and Prattichizzo, D. 2014. A remote guidance system for blind and visually impaired people via vibrotactile haptic feedback. *22nd Mediterranean Conference on Control and Automation* (Jun. 2014), 20–23.

[33] Schoof, Loren T., I. 1974. An analysis of Optacon usage. (Nov. 1974).

[34] Schwerdt, H.N., Tapson, J. and Etienne-Cummings, R. 2009. A color detection glove with haptic feedback for the visually disabled. *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on* (2009), 681–686.

[35] Sergi, F., Accoto, D., Campolo, D. and Guglielmelli, E. 2008. Forearm orientation guidance with a vibrotactile feedback bracelet: On the directionality of tactile motor communication. *IEEE/RAS-EMBS Biomedical Robotics and Biomechatronics, BioRob 2008*, 433–438.

[36] Shilkrot, R., Huber, J., Wong, M.E., Maes, P. and Nanayakkara, S. 2015. FingerReader: a wearable device to explore printed text on the go. *Proc. of CHI '15*, 2363–2372.

[37] Stearns, L., Du, R., Oh, U., Jou, C., Findlater, L., Ross, D.A. and Froehlich, J.E. 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *ACM Trans. on Accessible Computing*. 9, 1 (Oct. 2016), 1–38.

[38] Stearns, L., Du, R., Oh, U., Jou, C., Findlater, L., Ross, D.A. and Froehlich, J.E. 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *ACM Transactions on Accessible Computing*. (2016).

[39] Upton, G. and Cook, I. 1996. *Understanding statistics*. Oxford University Press.

[40] Weber, B., Schätzle, S., Hulin, T., Preusche, C., Deml, B., Schatzle, S., Hulin, T., Preusche, C. and Deml, B. 2011. Evaluation of a vibrotactile feedback device for spatial guidance. *IEEE World Haptics Conference'11*, 349–354.

[41] Wilson, G. and Brewster, S.A. 2016. Using Dynamic Audio Feedback to Support Peripersonal Reaching in Young Visually Impaired People. *Proc. of ASSETS'16*, 209–218.

[42] Wobbrock, J.O., Findlater, L., Gergle, D. and Higgins, J.J. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. *Proc. of CHI'11*, 143–146.

Recognizing Clothing Colors and Visual Textures Using a Finger-Mounted Camera: An Initial Investigation

Alexander J. Medeiros[1], Lee Stearns[1], Leah Findlater[2,5], Chuan Chen[3], Jon E. Froehlich[1,4]

[1]Department of Computer Science
[2]College of Information Studies
University of Maryland, College Park

[3]Poolesville High School
Poolesville, MD

[4]Computer Science and Engineering
[5]Human Centered Design and Engineering
University of Washington

ajmed@umd.edu, lstearns@umd.edu, leahkf@umd.edu, riverchen@gmail.com, jonf@cs.umd.edu

ABSTRACT

We investigate clothing color and visual texture recognition using images from a finger-mounted camera to support people with visual impairments. Our approach mitigates issues with distance and lighting that can impact the accuracy of existing color and texture recognizers and allows for easy touch-based interrogation to better understand clothing appearance. We classify image textures by combining two off-the-shelf techniques commonly used for object recognition achieving 99.4% accuracy on a dataset of 520 clothing images across 9 texture categories. We close with a discussion of potential applications, user evaluation plans, and open questions.

CCS Concepts

• **Human-centered computing** → **Accessibility technologies**

Keywords

Blind; visually impaired; wearables; texture recognition

1. INTRODUCTION

For people with visual impairments, color and texture information are frequently inaccessible. Standalone commercial devices such as *Color Teller* [2] or smartphone apps such as *Color Identifier* [5] allow users to identify colors and hear them read aloud. However, accuracy varies depending on ambient lighting and distance from the target surface. Also, these systems do not support recognition of visual textures or allow users to quickly query multiple locations—both of which are important to recognizing clothing [3]. Outside of commercial products, Yuan *et al.* [11, 13] investigated simple color and texture classification algorithms to identify clothing from static images. However, their approaches are again affected by distance and ambient lighting and do not allow for efficient interaction to identify more complex patterns.

In contrast, we introduce a wearable system using a small camera and co-located illumination source (an LED) worn on the finger. This system should enable access to color and texture information through touch, allowing users to move their finger across an article of clothing and combine knowledge of physical materials gained from their own sense of touch with automated audio feedback about visual appearance. This approach builds on a small but growing body of work, including our own [6, 10], that uses finger-mounted wearable cameras to support visually impaired users in a variety of activities. Our approach is most closely related to *Magic Finger* [12], which was not intended specifically for visually impaired users but which similarly uses a finger-mounted camera to classify touched surfaces including seven clothing textures. However, Magic Finger's classification approach was simplistic and would not scale well to a large database of textures. In contrast, our algorithmic approach is based on that by Cimpoi *et al.* [4], which combines two complementary features commonly used for object recognition to achieve state-of-the-art texture classification performance. To examine this approach with a finger-wearable camera, we conducted a classification experiment on a small custom dataset of 520 images across 9 visual texture categories and 29 articles of clothing, achieving 99.4% accuracy.

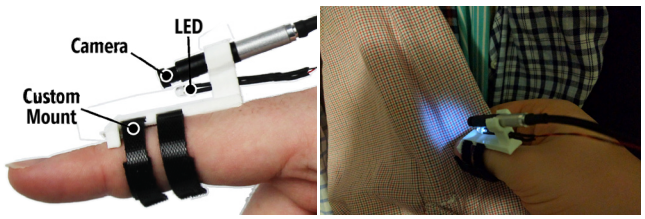


Figure 1. Prototype wearable camera and LED system: (a) Close-up view of system, (b) Identifying an article of clothing.

Our contributions include: (i) extending our finger mounted camera system [6, 10] to support robust color and visual texture recognition accessed through touch, (ii) a small but novel dataset of close-up clothing images captured using this system, and (iii) results of applying a state-of-the-art texture classification approach to our dataset. Because this work is at an early stage, we also cover open questions and challenges that will need to be addressed as we implement an interactive system for visually impaired users.

2. CAMERA AND ALGORITHMS

Our prototype system uses a small, endoscopic camera (*Awaiba NanEye GS Idule*) mounted on top of the finger to capture images of the touched surface (Figure 1). The camera provides 640×640px images with a 30º field of view and can focus from 15mm.

To identify textures, we use the classification approach described in [4], which combines two complementary features to improve classification performance. First, we detect deep convolutional activation features (DeCAF) using a pre-trained deep convolutional network. As in [4], we repurpose the *AlexNet* [7] image classifier for identifying textures by removing the last two softmax and fully connected layers. Second, we use scale-invariant feature transform (SIFT [8]) descriptors extracted densely at multiple scales. The SIFT features are combined into a single descriptor using the Improved Fisher Vector (IFV [9]) formulation. The result is two feature vectors of length 4,096 and 40,960 for DeCAF and IFV respectively, which are used as inputs (separately or concatenated together) to a support vector machine for classification.

We focus primarily on visual *texture* classification since few researchers have attempted to make this information accessible for visually impaired users. However, our *color* recognition approach

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ASSETS '17, October 29-November 1, 2017, Baltimore, MD, USA
© 2017 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-4926-0/17/10.
https://doi.org/10.1145/3132525.3134805

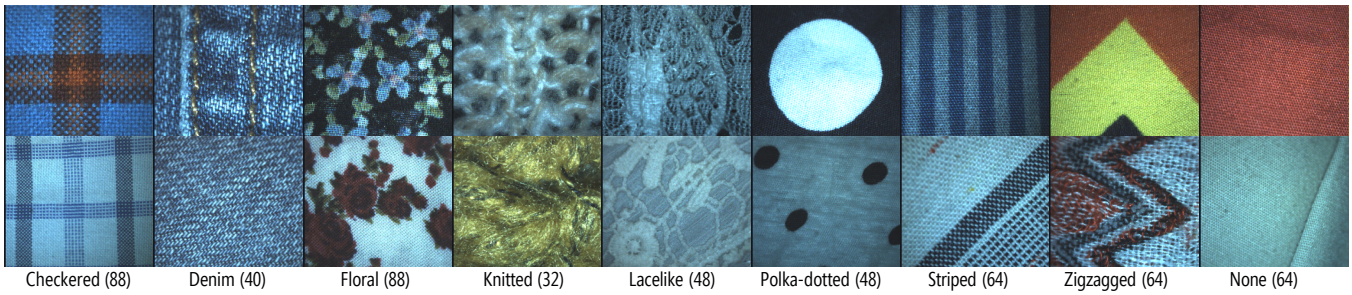


Figure 2. Examples of the 9 clothing textures included in our dataset. The numbers in parentheses indicate the quantity captured for each class.

also has some advantages over existing solutions. Our wearable system uses touch-based interactions to constrain the camera's distance from the target surface and includes a bright LED to overpower the effects of ambient light (allowing for more consistent color recognition). Furthermore, we use a color detection approach based on superpixel segmentation [1] that allows multiple colors to be detected simultaneously (one per superpixel) helping users to better understand how an article of clothing appears.

3. DATASET AND EXPERIMENTS

The results reported in [4] were promising but did not focus on clothing textures and used images from online sources that differed greatly from our target domain. It was unclear how well the approach would extend to close-up images captured by a finger-mounted camera. To answer this question, we collected a dataset of 520 images across 29 articles of clothing, which spanned 9 clothing texture categories (Figure 2). These categories are a subset of the 47 included in [4]; we eliminated categories that rarely describe clothing (*e.g.*, "bubbly"), combined those that are visually similar (*e.g.*, "striped", "banded), and added two new categories: "denim" and "none". We controlled for and varied the distance (5cm *vs.* 12cm), rotation (0° *vs.* 45°), and perspective of the camera (90° *vs.* 45°), as well as the tension of the fabric (taut *vs.* hanging naturally). The dataset was collected by one person over three months.

To assess performance, we conducted a classification experiment computing accuracy as the number of test samples classified correctly. We also explored the effect of training set size to determine whether a small user-gathered training set would be sufficient. Figure 3 shows classification accuracy using DeCAF and IFV features separately and together as the training set size increases averaged across 40 random samples.

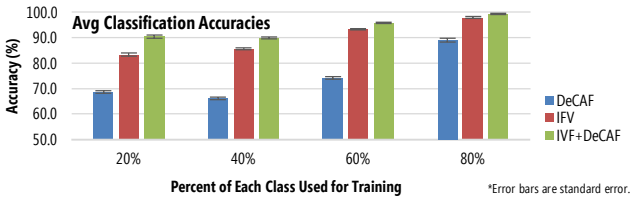


Figure 3. Accuracies using individual and combined features.

As with [4], in each case the combined result is best, demonstrating the advantage of using complementary texture features. However, unlike [4], our DeCAF results are significantly lower than IFV. This is likely because DeCAF requires a large amount of training data to perform well while IFV does well even with the small amount that we provided.

4. DISCUSSION AND OPEN QUESTIONS

Our results demonstrate the feasibility of recognizing clothing textures using close-up images from a finger-mounted camera. Even with a small amount of training data across a variety of variables, we achieve high classification accuracy. However, our dataset was captured under controlled conditions, so future work will need to investigate performance with a much larger, more varied dataset and conduct a user study with a real-time system.

We envision a few different scenarios in which visually impaired users could benefit from our proposed system. First, users could train the system to recognize customized articles of clothing in their own closet gathering training examples like those included in our dataset. Our results suggest reliable performance even with a small number of training examples. Second, we could draw upon larger texture datasets such as [4] to train a general-purpose classifier, which would allow users to access information about unfamiliar articles of clothing (*e.g.*, while shopping). Our approach should allow users to quickly explore a surface and combine their sense of touch with visual texture and color information to make informed decisions about what to wear or buy. How best to convey this information to visually impaired users is an open question.

5. ACKNOWLEDGMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

6. REFERENCES

[1] Achanta, R., *et al.*. 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *TPAMI*. 34, 11: 2274–2282.

[2] Brytech Color Teller: *http://www.brytech.com/colorteller/*.

[3] Burton, M.A., *et al.*. 2012. Crowdsourcing Subjective Fashion Advice Using VizWiz: Challenges and Opportunities. *Proc. ASSETS 2012*, 135–142.

[4] Cimpoi, M., *et al.*. 2014. Describing Textures in the Wild. *Proc. CVPR 2014*, 3606–3613.

[5] GreenGar Studios. Color Identifier: *https://itunes.apple.com/us/app/color-identifier/id363346987?mt=8*.

[6] Findlater, L., *et al.*. 2015. Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. *Proc. ASSETS* 2015, 383–384.

[7] Krizhevsky, A., *et al.*. 2012. ImageNet classification with deep convolutional neural networks. *Proc. NIPS 2012,* 1097–1105.

[8] Lowe, D.G. 2004. Distinctive image features from scale-invariant keypoints. *IJCV*. 60, 2 (2004), 91–110.

[9] Perronnin, F., Sánchez, J. and Mensink, T. 2010. Improving the Fisher Kernel for Large-Scale Image Classification. Springer, Berlin, Heidelberg. 143–156.

[10] Stearns, L., *et al.* 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *TACCESS*. 9, 1: 1–38.

[11] Tian, Y. and Yuan, S. 2010. Clothes Matching for Blind and Color Blind People. *Computers Helping People with Special Needs SE - 48*. Springer Berlin Heidelberg. 324–331.

[12] Yang, X.-D., *et al.*. 2012. Magic finger: always-available input through finger instrumentation. *Proc. UIST 2012*, 147–156.

[13] Yuan, S., *et al.*. 2011. Clothing matching for visually impaired persons. *Technology and Disability*. 23, 2 (Jan. 2011), 75–85.

Investigating Microinteractions for People with Visual Impairments and the Potential Role of On-Body Interaction

Uran Oh[1,2], Lee Stearns[1], Alisha Pradhan[3], Jon E. Froehlich[1,4], Leah Findlater[3,5]

[1]Department of Computer Science
[3]College of Information Studies
University of Maryland, College Park

[2]Robotics Institute
School of Computer Science
Carnegie Mellon University

[4]Computer Science and Engineering
[5]Human Centered Design and Engineering
University of Washington

uranoh@cmu.edu, lstearns@umd.edu, alisha93@terpmail.umd.edu, jonf@cs.umd.edu, leahkf@umd.edu

ABSTRACT

For screenreader users who are blind or visually impaired (VI), today's mobile devices, while reasonably *accessible*, are not necessarily *efficient*. This inefficiency may be especially problematic for microinteractions, which are brief but high-frequency interactions that take only a few seconds for sighted users to complete (*e.g.*, checking the weather or for new messages). One potential solution to support efficient non-visual microinteractions is on-body input, which appropriates the user's own body as the interaction medium. In this paper, we address two related research questions: *How well are microinteractions currently supported for VI users? How should on-body interaction be designed to best support microinteractions for this user group?* We conducted two studies: (1) an online survey to compare current microinteraction use between VI and sighted users (N=117); and (2) an in-person study where 12 VI screenreader users qualitatively evaluated a real-time on-body interaction system that provided three contrasting input designs. Our findings suggest that efficient microinteractions are not currently well-supported for VI users, at least using manual input, which highlights the need for new interaction approaches. On-body input offers this potential and the qualitative evaluation revealed tradeoffs with different on-body interaction techniques in terms of perceived efficiency, learnability, social acceptability, and ability to use on the go.

CCS Concepts

• Human-Centered Computing → Accessibility → Empirical Studies in Accessibility

Keywords

Visual impairments; mobile; microinteraction; wearable technology; on-body interaction.

1. INTRODUCTION

Today's mobile devices are reasonably *accessible* for blind and visually impaired (VI) users but are not necessarily *efficient*. This inefficiency may be especially problematic for microinteractions—that is, brief but high-frequency interactions that typically take sighted users only a few seconds to complete [3]. Manually finding and playing a song, for example, can take 15 seconds for a blind user [20] while entering a four-digit passcode to unlock a smartphone requires on average eight seconds, leading many blind users to forgo this security feature altogether [6].

Copyright block.

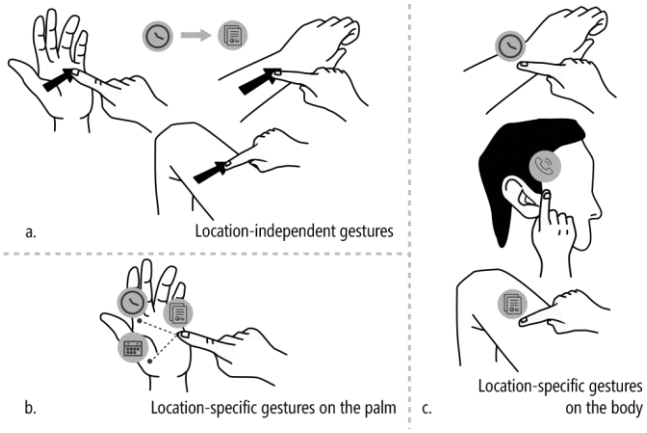


Figure 1. The three on-body interaction techniques explored for microinteractions in Study 2: (a) location-independent taps and swipes that can be performed anywhere on the body, (b) location-specific input that allows users to directly access a specific set of applications by tapping a dedicated location on the palm only or (c) on the body. See also video figure.

While many approaches could be taken to support efficient non-visual microinteractions, one particularly promising direction is *on-body input* [16,17], which appropriates the user's own body as the interaction medium. Prior work has shown that on-body input can be more efficient than touchscreen interaction for non-visual use, even without including the time to first retrieve the smartphone from a pocket or bag [14,26].

In this paper, we address two interrelated research questions: *(1) How well are microinteractions currently supported for blind and visually impaired users? (2) How should on-body interaction be designed to best support microinteractions for this user group?* Prior work in on-body interaction for users who are blind or visually impaired has identified a desire for socially acceptable input locations and the potential to support use on-the-go [25]. However, important issues related to efficient input remain, including understanding VI users' experiences with existing microinteractions (*e.g.*, on smartwatches, smartphones) and reactions to on-body interaction, as well as how to best leverage the body for microinteraction input (*e.g.*, taps and swipes anywhere on the body *vs.* at specific locations; Figure 1).

To investigate current support for accessible microinteractions and the design of on-body microinteractions for VI users, we conducted two studies. The first study was a formative online survey with 117 users (56 sighted and 61 VI) to characterize current microinteraction use and identify barriers therein. For the second

study, we recruited 12 screenreader users who participated in an in-person semi-structured interview on the same themes and completed a qualitative evaluation of a real-time on-body interaction prototype that we developed. For the latter, participants compared three complementary on-body interaction techniques: location-independent gestures (*i.e.*, taps and swipes anywhere on the body), location-specific gestures on the hand (*i.e.*, pointing to different hand locations), and location-specific gestures on the body (*i.e.*, pointing to different hand and body locations).

Our findings suggest that efficient microinteractions are not currently well-supported for VI users, at least using manual input. While speech input is a common alternative, it is not always appropriate, thus highlighting the need for new approaches to support frequent and fast non-visual interactions. The qualitative evaluation of the three on-body interaction techniques to support microinteractions revealed tradeoffs in terms of perceived efficiency, learnability, social acceptability, and the ability to use on the go. For example, participants considered input at *specific* locations across the body to be the least efficient whereas location-independent gestures, which can be performed anywhere, were seen as supporting learning and offering flexibility for use on-the-go. In summary, the primary contributions of this paper are: (1) an identification of the needs, barriers, and strategies for enabling microinteractions for users with visual impairments in comparison to sighted users, and (2) design implications for supporting microinteractions for VI users via on-body input.

2. RELATED WORK

Our research is informed by prior work on microinteractions and mobile and wearable technologies for VI users.

2.1 Microinteractions

Microinteractions are defined by Ashbrook [3] as interactions that take less than four seconds to complete, such as changing the volume or dialing a number. These interaction durations are important because they can profoundly influence usage of an application or device [28,34]. Mobile contexts also increase the need for efficient interaction: Oulasvirta *et al.* [28] showed that mobile interaction typically occurs in short bursts of attention that are only four to eight seconds long.

Research on understanding and supporting microinteractions has largely focused on sighted users. Pizza *et al.* [29] found that the most frequent smartwatch interactions under 10 seconds were glancing at the watch face, accessing notifications, and setting and checking timers. As well, in an analysis of over 4,000 Android smartphone users, Bömer *et al.* [7], found that tasks related to communication (*e.g.*, text messaging, phone calls) were among the shortest and most frequently used tasks. Efforts to support more efficient microinteractions have largely focused on wearable devices because wearables enable quicker access times and less disruptive transitions between the device and the physical world compared to mobile phones [4,22,24]. In comparison to sighted users, however, VI users—our focus—typically have longer interaction times [5,6], which is particularly problematic in mobile contexts since device access is already limited [1,33,39]. While some wearable devices for VI users arguably support microinteractions (e.g., *EyeRing*'s barcode scanning and currency identification [24]), to our knowledge, no study has thoroughly examined VI users' microinteraction needs and how well these needs are currently supported.

2.2 Accessible Mobile/Wearable Interaction

Mobile technology plays an important role for VI users, particularly in supporting independence [21]. Yet, touchscreen smartphones are not innately accessible due to a heavy reliance on visual cues and the minimal tactile feedback of the smooth touchscreen [18,23]. One approach for non-visual use is to employ location-insensitive gestures that do not require accurate hand-eye coordination [8,13,20]. For example, *Apple VoiceOver*[1] and *Google TalkBack*[2], two popular mobile screenreaders, provide location-insensitive taps and swipes that can be performed anywhere on the screen. These solutions also support location-specific interaction by allowing users to touch down and move the finger around to hear audio output for each element that is touched. Despite these advances, however, touchscreen interactions can still be time-consuming for VI users [5,6,20,39]—likely impacting microinteraction efficiency.

Accessible wearable interaction has largely focused on supporting tasks in the physical world, such as way-finding [11], object or character recognition [32,35], and currency detection [24]. For instance, *OrCam* includes a wearable camera and provides speech output for reading or identification when the user points to a document, object, or face [27]. A smaller body of work has focused on providing mobile digital information access through wearable technologies for VI users, such as Ye *et al.*'s [39] study of a wristband that pairs with a phone to control audio screenreader output. That study also investigated the advantages and limitations of wearable devices over smartphones for VI users and found that potential advantages included quick interaction, discretion, and use on the go. We build on these prior studies by focusing on microinteractions specifically and the ability of on-body input to support these interactions for VI users.

On-body input [16,17] offers the same advantages as most other wearable input: portable and quick to access. Moreover, with tactile and proprioceptive feedback from the user's own body, on-body input is efficient for eyes-free use [14]—a finding that has been extended to VI users [26]. Studies on the *design* of on-body interaction have focused primarily on sighted users (*e.g.*, [15,38]). For example, Weigel *et al.* [38] investigated characteristics of different skin input modalities and preferred input locations, and found that the most preferred location was the palm. As an exception, Oh and Findlater [25] studied the design of on-body input by asking VI users to create on-body input gestures and to compare location-independent input gestures on the hand to on a touchscreen phone. Findings showed that input on the hands and arms was considered relatively socially acceptable and on-body input in general was seen as valuable compared to the phone during hands-busy use (*e.g.*, one hand holding a cane or dog leash). However, the authors did not investigate *location-specific* on-body input, which has been shown to offer potential performance benefits over touchscreen interaction [26].

3. STUDY 1: MICROINTERACTION SURVEY

We conducted an online survey with 61 VI and 56 sighted smartphone users to compare microinteraction usage patterns between these two groups. We targeted VI users who interact primarily through audio output (*e.g.*, screenreaders) rather than visual output (*e.g.*, screen magnifiers). Thus, our VI group only includes participants who reported using screenreaders most or all of the time when interacting with their mobile device.

[1] http://www.apple.com/accessibility/ios/voiceover/

[2] https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback

Smartwatch: A smartwatch offers many of the same features as a smartphone, but has a smaller screen that is roughly the size of a large watch face. Smartwatches often include a camera, speaker, microphone, and sensors that can track information like the number of steps you've taken.

On-body interaction: Imagine a small wearable device such as a wristband or a ring that can sense when you do taps, swipes or other gestures on the surface of your body. This device can be paired with a small speaker for audio output or with a projected image on your arm or hand and would allow you to do many of the same actions as you can do on a smartphone or smartwatch. For example, you can perform taps and swipes on your bare palm in the same way you usually use the touchscreen on a phone or smartwatch. You can also tap or perform other gestures at specific locations on your body, such as the wrist, a fingertip, or ear to execute specific actions (e.g., check the time, answer a phone call, change a song's volume).

Figure 2. Descriptions of smartwatches and on-body interaction used in the online survey.

The survey focused on microinteractions with smartphones and smartwatches but also solicited open-ended feedback about on-body interaction. We hoped to learn:

1. What are the most common microinteractions currently used by sighted and VI users?
2. What tasks, if any, are not supported as microinteractions for VI users but would be valuable to support in the future?
3. How do users perceive the use of wearable devices (specifically, smartwatches) for microinteractions compared to smartphones?
4. How do people react to the idea of on-body interaction and what potential use cases do they foresee?

3.1 Method

We created an accessible online survey via *SurveyMonkey*, which was designed to take ~25 minutes for screenreader users. Smartphone owners were recruited to participate through email lists, a university bulletin board, community organizations, Facebook, and word of mouth. After completion, participants could opt into a drawing for a $100 Amazon gift certificate.

3.1.1 Survey Design

The survey, included in Appendix A, consisted of 32-36 questions depending on the participant's level of vision. For smartwatch owners, we asked an additional 24 questions. Questions included general background (*e.g.*, age, gender), current mobile and wearable technology use (*e.g.*, device type, frequency of use), and perceived tradeoffs between smartphones and smartwatches. The survey also asked respondents to estimate the time needed for them to complete specific microinteractions that are common for sighted users on mobile and wearable devices [2,37]—see Table 1. At the end of the survey, we also solicited feedback on the idea of on-body interaction. To aid understanding, we provided brief descriptions of smartwatches and on-body interaction (Figure 2).

3.1.2 Data and Analysis

During a one-month period, we received 134 fully completed and 13 partially completed survey responses (dropout rate of 8.8%). Fifty-six participants reported normal or corrected-to-normal vision (*sighted* participants) and 61 participants reported having a visual impairment and using a screenreader most or all of the time (*VI* participants); an additional 17 VI participants who were infrequent screenreader users were excluded from analysis. Of the VI group, one participant had 20/70 to 20/200 vision, 13 were legally blind (at best 20/200), 17 were blind with some light perception, and 30 were totally blind. Participants were asked to specify their age within a range (*e.g.*, 35–44, 45–54); the median age range was 45–54 for VI participants (38 female, 22 male, 1 other) and 25–34 for sighted participants (30 female, 26 male).

Table 1. The ten microinteractions examined in our online survey. This list is based on [7,29].

Task Label	Task Description
Alarm	Set an alarm or timer
App launch	Find and open a specific app
Calendar	Check your calendar for an overview of the day's schedule
Clock	Check the current time
Music	Pause a music player
Navigation	Set a destination to get navigation directions
Phone calls	Dial a phone number
Read message	Read a text message that is two sentences long
Respond to msg.	Respond to a text message with a two-word reply
Weather	Check the weather

Table 2. Mobile tasks reported as being frequent and fast to complete (≤ 10 seconds) by at least five VI or sighted participants, sorted by popularity among VI participants. Bolding and '*' indicate tasks that were significantly different between the two user groups.

Task	VI (*N*=61)	Sighted(*N*=56)
Email interactions, primarily checking for new email	80.3%	73.2%
Text message interactions, primarily checking for new messages	68.9%	76.8%
Checking the weather*	41.0%	21.4%
Voice call interactions (*e.g.,* answering or declining calls, dialing, checking missed calls)*	34.4%	10.7%
Checking the time	31.1%	28.6%
Checking notifications other than calls, emails, text messages, social media	26.2%	19.6%
Social media, primarily checking for updates	19.7%	30.4%
Information search (*e.g.,* trivia, bus schedule)	19.7%	17.9%
Music player control	18.0%	8.9%
Changing device settings (*e.g.,* volume, wifi)	9.8%	5.4%
Calendar interactions	8.2%	8.9%

We used chi-square tests of independence to analyze the impact of user groups (VI *vs.* sighted) on frequency data (*e.g.*, from multiple-choice questions on microinteraction usage). For the seven open-ended questions in the survey, we followed an iterative coding process [19]. Two researchers together developed initial codebooks for each question, then independently coded a randomly selected subset of 20-40 responses. Cohen's kappa was computed to assess interrater reliability for each code and problematic codes were refined; three to four iterations of this process were completed for each question. The average kappa score across all codes after the final iteration was 0.87 (*SD* = 0.12; range 0.53–1.0).

3.2 Findings

We describe participants' experiences with microinteractions, perceived tradeoffs between smartphones and smartwatches, and responses to the idea of on-body interaction. Asterisks in tables and figures indicate statistically significant differences ($p < .05$) between the VI and sighted participant groups.

3.2.1 Device Ownership and Frequency of Use

As required to complete the survey, all participants owned smartphones. Most VI participants had iOS devices (59 iOS, 8 Android, 1 other), while sighted participants were more evenly split (29 iOS, 30 Android, 5 other), confirming previously identified patterns of adoption [5,37,39]; some participants reported multiple devices. Participants used their devices frequently: 56 VI and 55 sighted participants reported using their smartphone at least once every few hours. Thirteen participants (4 VI, 9 sighted) owned a smartwatch and all reported using it at least every few hours.

3.2.2 Frequently Used Microinteractions

To compare what current smartphone tasks can be classified as microinteractions for VI users versus sighted users, we asked

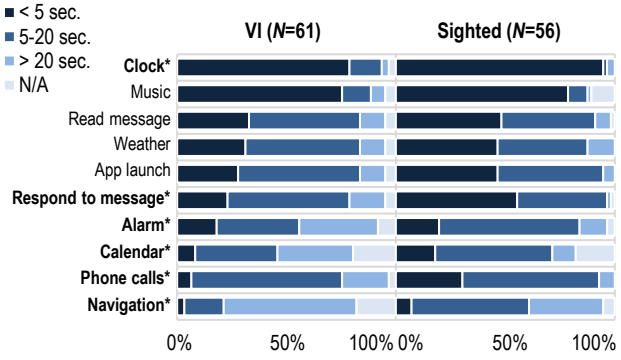


Figure 3. Percentage of survey participants reporting different task completion times across 10 common microinteraction tasks, sorted by the percentage of VI participants who reported *'< 5 seconds'*.

participants to list tasks that were *frequent* and *fast* (≤ 10 seconds to complete). This was an open-ended question and responses are shown in Table 2. While overall trends are similar across the two groups, VI participants were significantly more likely to list *'checking the weather'* (41% *vs.* 21%; $\chi^2_{(1)} = 5.16$, $p = .023$, $\phi = .21$) and *'voice call interactions'* (34% *vs.* 11%; $\chi^2_{(1)} = 9.25$, $p = .002$, $\phi = .28$) than sighted participants. The voice call interactions by VI users consisted primarily of checking for missed calls and voicemails. These differences may reflect the utility of auditory information for VI users, such as hearing a weather forecast versus looking out the window or talking versus composing a text message. As for accessibility-specific tasks such as OCR or identifying colors, only one participant reported such a task as taking 10 seconds or less to complete: '*identifying currency'*.

3.2.3 Task Completion Time for Microinteractions

To quantify how screenreader use impacts microinteraction task completion time, we asked participants to estimate their time to complete each of the 10 microinteractions listed in Table 1 using manual input (*i.e.*, without speech dictation); options were *'< 5 seconds'*, *'5-20 seconds'* and *'> 20 seconds'*. Figure 3 shows the reported durations for each task. Generally, a greater percentage of sighted participants reported being able to complete the tasks in *'< 5 seconds'* compared to VI users. The frequencies of reported completion times were significantly different between the two participant groups for more than half of the tasks (Table 3). While prior work has shown that screenreader users are slower at mobile tasks than sighted users [5,21], these results emphasize how widespread this trend is for tasks that are often deemed to be microinteractions. Focusing on the responses that were *'< 5 seconds'* (a microinteraction duration), the most substantial difference between the two groups was related to text entry (*'Respond to Message'*). Only 23% of VI participants reported *'< 5 seconds'* for this task compared to 55.4% of sighted participants, perhaps due to the inefficiency of manual text entry for users with visual impairments [5,8].

3.2.4 Use of Speech Input

Motivated by studies showing that speech input (*e.g.*, *Apple Siri*) is more popular for blind users than sighted users [5,39], we examined speech input use for microinteractions. Overall, and confirming prior work, a significantly greater number of VI participants (36.1%) reported using speech input at least '*most of the time*' compared to sighted participants (8.9%); chi-square test ($\chi^2_{(1)} = 12.11$, $p = .001$, $\phi = .32$). We also asked participants to indicate whether they '*usually'* use speech input for each microinteraction

Table 3. Chi-square test results comparing reported microinteraction task durations between the sighted and VI groups (does not include responses of "N/A"). Only the six tasks (of 10) that were significantly different are listed.

Task	N_{VI}	$N_{Sighted}$	$\chi^2_{(2)}$	p	ϕ
Clock*	59	56	6.57	.038	.24
Respond to Message*	58	55	15.84	<.001	.38
Phone Calls*	59	56	13.38	.001	.34
Alarm*	56	54	10.59	.005	.31
Calendar*	49	46	10.84	.004	.34
Navigation*	50	53	15.18	.001	.39

Table 4. Percentage of participants who *'usually'* use speech input for each of ten given microinteractions, with chi-square test results comparing the two participant groups. Bolding and '*' indicate tasks that were significantly different between the two user groups.

Task	VI (*N*=61)	Sighted (*N*=55)	$\chi^2_{(3)}$	p	ϕ
Phone calls*	73.8%	25.0%	27.78	<.001	.49
Alarm*	72.1%	33.9%	17.15	<.001	.38
Respond to message*	67.2%	17.9%	28.93	<.001	.50
Navigation*	63.9%	30.4%	6.73	.01	.24
Weather*	52.5%	21.4%	11.98	.001	.32
App launch*	50.8%	16.1%	15.67	<.001	.37
Read message*	34.4%	7.1%	12.94	<.001	.33
Clock*	34.4%	7.1%	12.94	<.001	.33
Calendar*	32.8%	7.1%	11.78	.001	.32
Music	16.4%	7.1%	2.37	.124	.14

listed in Table 1 (responses: yes/no). VI participants were significantly more likely to use speech input than sighted participants for all tasks except '*Music*' (Table 4). The biggest effect size was for '*Respond to Message*' where VI users were almost 4x as likely to use speech, again perhaps reflecting the difficulty of manual text entry for this user group [5,8].

3.2.5 Comparison of Smartphones and Smartwatches

While wearable devices like smartwatches have been shown to be particularly useful for supporting microinteractions for sighted users [3], we wanted to understand whether perceptions differ for VI users. Specifically, we asked four open-ended questions comparing smartwatches and smartphones: two about perceived advantages/disadvantages of smartwatches (Table 5) and two about tasks conducive to each technology. Because only a small number of participants owned a smartwatch (9 sighted, 4 VI), our analysis does not just focus on their responses but instead includes all 117 participants.

Table 5 shows smartwatch advantages and disadvantages mentioned by at least five participants in either user group, along with chi-square test results comparing the groups. VI participants were significantly more likely to mention advantages due to the watch's small size (*e.g.*, lightweight, portable) (30.9% *vs.* 12.7%), while sighted participants were significantly more likely to mention disadvantages due to size (*e.g.*, small input/output area) (72.7% *vs.* 53.7%). VI participants were also significantly more concerned about '*sound-related issues*' such as low volume or not being able to pair with/plug in headphones (16.7% *vs.* 0.0%). Overall, these differences may reflect a reliance on visual output by sighted users compared to audio output by VI users.

When participants were asked about tasks they would prefer to do on a smartwatch compared to a smartphone, the top responses were fitness tracking (39.3% of VI *vs.* 33.9% of sighted participants), clock (32.8% *vs.* 37.5%), and text messaging (32.8% *vs.* 33.9%). Interestingly, 29.3% of VI participants said they would prefer the

Table 5. Perceived advantages and disadvantages of smartwatches compared to smartphones mentioned by at least five VI or sighted participants, sorted by popularity among VI participants. Items with significantly different response rates between the two groups are bolded and marked with '*'.

Advantages	VI (*N*=55)	Sighted (*N*=55)	$\chi^2_{(1)}$	*p*	ϕ
Quick/easy access	50.9%	52.7%	0.04	.849	-.02
Small – lightweight*	**30.9%**	**12.7%**	**5.33**	**.021**	**.22**
Portable	21.8%	14.5%	0.98	.323	.09
Not having to hold a device	9.1%	9.1%	0.00	1.00	.00
Less distracting*	**0.0%**	**9.1%**	**5.24**	**.022**	**-.22**

Disadvantages	VI (*N*=54)	Sighted (*N*=55)	$\chi^2_{(1)}$	*p*	ϕ
Small – input/output space*	**53.7%**	**72.7%**	**4.25**	**.039**	**-.20**
Limited functionality	22.2%	16.4%	0.60	.438	.07
Sound-related issues*	**16.7%**	**0.0%**	**9.99**	**.002**	**.30**
Not a stand-alone device	14.8%	10.9%	0.37	.542	.06

phone for all tasks compared to only 13.2% of sighted participants; this difference was significant ($\chi^2_{(1)} = 4.24$, $p = .040$, $\phi = .20$).

3.2.6 Attitude Toward On-body Interaction.

Finally, we asked two open-ended questions about on-body interaction: overall reactions to the concept and for what specific tasks, if any, such interaction may be useful. Five VI participants did not answer this question and are excluded from this analysis. Participants were roughly evenly split in mentioning potential benefits or concerns of on-body interaction. 48.2% of VI and 60.7% sighted participants mentioned at least one benefit, such as quick/easy access or larger interaction surface than a smartphone. However, 46.4% of VI and 46.4% of sighted participants also mentioned at least one concern, such as social acceptability, technical challenges (*e.g.*, sensing accuracy), and learning curve. The most frequent concern for VI users was social acceptability, although this was mentioned by only 12.5% of VI participants; 3.6% of sighted participants reported the same. Fifty-five VI users and 51 sighted users provided suggestions for tasks where on-body interaction would be especially useful. The most common were making phone calls (29.1% of VI and 39.2% of sighted responses) and checking the time (27.3% of VI and 31.4% of sighted responses). Other suggestions included changing device settings, reading/checking new messages, and controlling a media player.

3.3 Summary

Our findings emphasize differences between VI and sighted users' experiences with microinteractions. VI users took longer to complete microinteractions for most tasks and were more likely to use speech input. In terms of attitudes towards smartwatches, VI users were more likely to appreciate the small size of the device but to also be concerned about audio quality compared to sighted users. Both groups of participants responded largely similarly to the idea of on-body interaction, expressing potential benefits (*e.g.*, larger interaction space) and some concerns (*e.g.*, social acceptability).

4. Study 2: Interview & Prototype Evaluation

We conducted an in-person study with 12 VI participants to re-examine the survey themes in more depth and to subjectively compare different on-body interaction techniques for microinteractions. Previous work has shown that on-body input is more efficient for VI users than smartphone input when using *location-specific* gestures—such as pointing to different areas of the hand [26]. But, how does such input compare subjectively to more flexible *location-independent* input, such as taps or swipes anywhere on the body? Moreover, for location-specific input, do users value semantic mappings between locations and microinteractions (*e.g.*, tapping the wrist to hear the time) or more discreet but less meaningful locations? To address these questions, we built and asked participants to use a real-time system that supported three on-body interaction techniques: location-independent input, location-specific input at different body locations, and location-specific input on only the hand.

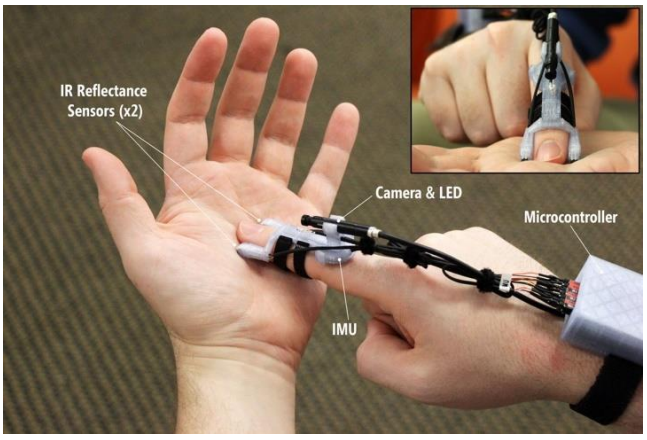


Figure 4. Our wearable, on-body interaction prototype system consists of a finger-worn package including a camera and LED, two IR reflectance sensors, and an IMU. Computer vision and machine learning algorithms are used to support location-specific, on-body gestural interaction.

4.1 Participants

We recruited 12 smartphone users with visual impairments (7 female) through email lists, local organizations, and word of mouth; three of them (P1, P6 and P9) also participated in Study 1. Nine were blind and three had low vision. They were on average 46.2 years old (*SD* = 12.0, range 29–65). All participants had owned a smartphone for more than a year and used a screenreader "*all of the time*", except for P11 who used a screenreader "*most of the time*." P5 used an Android phone, while the remaining participants used iPhones. Only two participants owned a wearable device: P10 had a fitness tracker and P5 had previously owned a smartwatch but rarely used it. As with the survey, the use of speech input was prevalent: ten participants used speech at least some of the time; seven of these reported regular use. Participants were compensated $60 for their time and travel costs.

4.2 Real-time Wearable System

We built a real-time wearable system, shown in Figure 4, to explore microinteractions using three different on-body interaction techniques: location-independent gestures, location-specific gestures on the palm only, and location-specific gestures at a wider variety of body locations.

4.2.1 System and Algorithms

The system includes several finger-worn sensors, as shown in Figure 4: a pair of infrared (IR) reflectance sensors[3] near the tip of the finger to sense *touch-down* and *touch-up* events, a camera[4] that localizes the user's touch using computer vision, an LED to provide consistent lighting for the camera, and an inertial measurement unit (IMU)[5] for sensing movements and recognizing gestures. Placing the sensors on the gesturing finger rather than elsewhere on the

[3] Fairchild Semiconductor QRE113GR
[4] Awaiba NanEye GS Idule Demo Kit
[5] Adafruit Flora LSM9DS0

Table 6. Our system supports 16 microinteractions organized in a two-level hierarchical menu. An * identifies items that we asked participants to select in our user study.

Category	Application	Description
Clock	Time	Check the current time
	Alarm*	Check the next alarm
	Timer*	Check the time remaining
	Stopwatch	Check the time elapsed
Daily Summary	Date	Check today's date
	Calendar*	Check the next event
	Weather*	Check the current temperature and weather conditions
Notifications	Summary	Summarize notifications (*e.g.*, "one missed phone call and two new messages")
	Missed phone call*	Check missed phone calls (*e.g.*, "missed phone call from Alice")
	New message #1	Check first new message (*e.g.*, "message from Bob")
	New message #2*	Check second new message (*e.g.*, "message from Charlie")
Health and Activities	Distance	Check the miles traveled
	Steps*	Check the number of steps taken
	Calories	Check the calories burnt
	Heart rate*	Check the heart rate
*Speech Input**	(No sub-menu)	Activate voice input

body mitigates the camera framing issues that VI users often face [2] and enabled flexible input at a variety of locations.

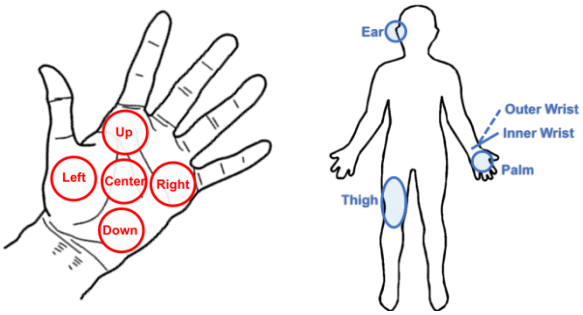
Our system use the camera images that occur between touch-down and touch-up events to localize patches of skin, an approach adapted from [36]. While the algorithms are not the focus of the current paper, in short, the approach is as follows. The visual texture, represented as a histogram of rotation-invariant local binary patterns at multiple scales, is classified using support vector machines (SVMs). To achieve real-time use (~35 fps), we remove the geometric verification stage reported in [36] and compensate for the resulting loss in accuracy by combining location estimates across 20 video frames (~570ms window). Once the location has been classified, an SVM trained on accelerometer, gyroscope, and IR features classifies the gesture itself (*e.g.*, tap, directional swipe). The image-based localization step requires training per participant, as described in the procedure section below.

The hardware components were mounted on the user's index finger via a 3D-printed ring and Velcro strips, and were positioned to avoid interfering with the user's fingertip sensitivity or range of motion. The IMU and IR sensors were connected to a wristband containing an Arduino microcontroller[6], which in turn connected along with the camera to a desktop computer[7]. Speech feedback was provided through a pair of speakers using the *Microsoft .NET* speech synthesis libraries. This prototype was intended to explore possible on-body interaction designs, and we envision a future system to be much smaller and self-contained.

4.2.2 Interaction Techniques

The system allowed users to access 16 applications that align with the microinteraction findings in Study 1. As shown in Table 6, the applications were arranged in a two-level menu hierarchy with five top-level categories; only the *Speech Input* category had no sub-menu. We implemented the following three interaction techniques, shown in Figure 1 and in the accompanying video figure, for accessing the top-level categories.

Location-independent gestures (LI). Touching anywhere on the body and swiping left and right allows the user to linearly navigate through the top-level menu categories, while a double-tap activates



(a) Five LS_{palm} locations **(b) Five LS_{body} locations**

Figure 5. Locations used for LS_{palm} and LS_{body} interaction techniques: (a) LS_{palm} mapped five microinteraction categories arbitrarily to five locations on palm; (b) LS_{body} mapped the categories when possible to semantically meaningful body locations (e.g., wrist for *Clock*).

the currently selected category. This interface is simple and flexible, but similar location-independent interaction on the touchscreen has been shown to be time-consuming for VI users navigating a long list of items [20].

Location-specific gestures on the palm (LS_{palm}). The top-level categories are arbitrarily mapped to the five palm locations shown in Figure 5a: up (*Notifications*), left (*Clock*), right (*Daily Summary*), down (*Health and Activities*), and center (*Voice Input*). Users can select a category by directly tapping on its location or they can "touch and explore"—touching down anywhere on the palm, sliding the finger around to hear each touched item read aloud, then lifting up when the desired item is found. VoiceOver and TalkBack support a similar touch-and-explore functionality.

Location-specific gestures on the body (LS_{body}). The top-level categories are mapped to five locations on the body, with semantic mappings where possible (Figure 5b). Semantic body mappings, have shown promise for body-centric mobile interactions [10,12]. The five locations included: tapping the outer wrist for *Clock*, the ear for *Speech Input*, and the thigh for *Health and Activities* (*e.g.*, step count). Other mappings were the palm for *Notifications* and the inner wrist for *Daily Summary*.

Once the top-level category is selected and activated using one of these three interaction techniques, users swipe left/right to navigate the submenu items and double-tap to select the current item. This submenu interaction is similar to basic interface navigation with VoiceOver and TalkBack, with which our participants were familiar (all were screenreader users). Of the three interaction techniques, the location-specific options should theoretically be fastest because they allow for direct access to menu categories, while the location-independent option provides more flexibility.

4.3 Procedure

The study session began with a background questionnaire, followed by semi-structured interview questions (~20 minutes) that focused on the same themes as the survey. After the interview, we calibrated and trained the localization component of the wearable on-body interaction prototype (~30 minutes), which included collecting several images of each location shown in Figure 5. The three interaction techniques were then presented in counterbalanced order. For each interaction technique, participants were given a short introduction (~5 minutes) before independently completing a set of 10 microinteractions. Each microinteraction consisted of navigating to, selecting, and briefly using an application (*e.g.*,

[6] Sparkfun Arduino Pro Micro (5V/16MHz)

[7] Dell Precision T7910 (Intel Xeon, 8-core, 2.1Ghz, NVIDIA GeForce GTX 750Ti)

Alarm item within the *Clock* category). The set included two applications from each of the five menu categories (Table 6) except for the *Speech Input* category, which was simply selected twice since it had no submenu items. For each microinteraction, the tester stated which application to select along with its category. The session concluded with open-ended questions about the user experience and comparison of the three interaction techniques in terms of preference, ease of use, efficiency, and use in public.

4.4 Data Analysis

We audio-recorded and transcribed participants' open-ended responses for analysis. For questions that were also in the online survey, we reused the same analytical codebook (*e.g.*, "*What tasks or actions do you do frequently on your phone?*"). Other questions were analyzed by one researcher based on themes of interest [9] (*e.g.*, one-handed use of phone, interaction on-the-go), while allowing for new, emergent themes.

4.5 Findings

We briefly cover findings that confirm and extend the survey responses on general microinteraction use before presenting subjective responses to the on-body interaction techniques.

4.5.1 Current Microinteraction Experience

In general, trends were similar to the survey responses from VI participants, thus we highlight only new findings and examples.

Overall microinteraction use. The most frequently used and fast (≤ 10 seconds) smartphone tasks with manual input were: voice call interactions (*N*=9 participants out of 12), text message interactions (*N*=8), looking up information (*N*=6), and checking the weather (*N*=4). Seven participants specified that some tasks take less than 10 seconds only when using speech input (*e.g.*, voice search, dictation) emphasizing again the importance of speech input for supporting efficient interaction for VI users. In terms of accessibility-related tasks, three participants mentioned frequent use of *TapTapSee*[8], a money/object identification app, although it was not considered fast (≤ 10 seconds).

Microinteractions on-the-go. Previous work on the design of on-body interaction for VI users emphasized the importance of one-handed interaction to support use on-the-go, since users may be holding a cane or service dog leash [25]. Participant responses confirm that finding. When asked about using their phone with one hand, eight participants commented that one-handed use is important when walking. At the same time, nine participants also preferred not to use their phone at all when using a cane due to safety concerns, confirming that the perceived cost of interacting with a mobile device on-the-go is high for VI users [1,33,39]. For example, P8 said, *"See, when I'm walking, I have to use my cane, so I don't even be [sic] on the phone. To me, that's almost like driving and being on the phone. […] That'll put a blind person in danger."* Although limited to answering/rejecting a call or using speech input, six participants reported that they use a headphone or a headset to remotely control their phone while walking.

Advantages and limitations of smartwatches. We also asked about perceived advantages and limitations of smartwatches compared to smartphones. In general, responses reflected trends seen in the survey, such as advantages of quick/easy access (*N*=6), hands-free (*N*=5), and portability (*N*=4), and limitations related to the small size (*N*=9), functionality (*N*=3), need for a paired phone (*N*=3), and sound-related issues (*N*=2). Of the nine participants who reported limitations related to size, eight specified that the small screen would make input difficult, for example:

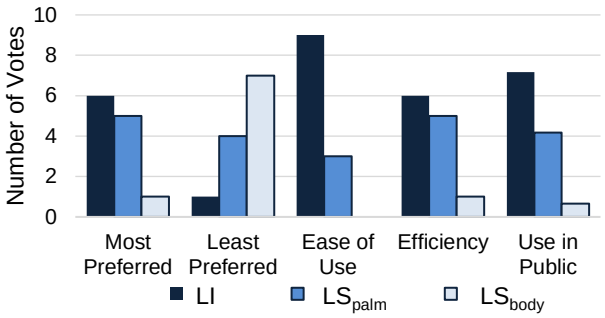


Figure 6. Vote counts for the interaction techniques that are the most and least preferred overall, most efficient, easiest to use, and most comfortable to use in public. Participants were allowed to vote for multiple options for *Use in Public*, but in such cases the votes were adjusted to sum to 1.0 per participant. (*N* = 12)

> *"…because [the phone] has [a] wider space so I can find the layout of the keyboard, and I'll have a wider area of navigation between the letters and the keys. But in the case of [a] smartwatch, it's much, much smaller, so it will be compact and difficult to locate each letter."* (P1)

4.5.2 On-body Prototype Evaluation

We summarize overall reactions to on-body interaction then compare subjective responses to our three interaction techniques.

General reactions. We asked participants for their thoughts about on-body interaction *before* using our prototype and again *after* using it in comparison to smartphones and smartwatches (open-ended). Most participants (*N*=9) reacted positively beforehand, citing similar reasons as in the survey: quick and easy access (*N*=7), being able to map different body locations to different interactive tasks (*N*=6), and not having to hold the phone in hand (*N*=4). After use, when compared with a smartwatch or a smartphone, reactions focused on technical/physical aspects of the interaction, such as the number of devices to be carried (*N*=6) or sensor/speaker locations (*N*=6). After using the prototype, three participants still felt that they would prefer to use their phone due to familiarity. For instance, P9 said: *"I would rather use the phone. […] I have no problem with using the body. It's just that I'm not used to it."*

Preferred interaction techniques. Figure 6 shows the subjective votes along several dimensions: overall most and least preferred, easiest to use, most efficient, and most appropriate for public use. Due to the sample size (*N*=12), we do not report on statistical tests on this data, but instead analyze the subjective rationale provided to support the votes. For overall preference, location-independent gestures (*LI*), where taps and swipes could be performed at any location, and gestures performed at specific locations on the palm only (LS_{palm}) were preferred to gestures performed at specific locations across the body (LS_{body}). The reasons for these preferences are broken down in the sections below.

Ease of use and learnability. As shown in Figure 5, most participants (*N*=9) felt that *LI* was the easiest to use of the three interaction techniques, for reasons such as: the microinteractions were easy to perform (*N*=7), the interaction was similar to that on a phone (*N*=6), and the input location was flexible (*N*=5). At the same time, about half of the participants considered the need to learn and remember specific locations to be a drawback of both LS_{palm} (*N*=7)

[8] http://www.taptapsee.com

and LS_{body} (*N*=6), although the semantic mappings of LS_{body} (*e.g.*, tap wrist for *Clock*) were seen by some (*N*=5) as beneficial for learning. Another valued interaction was the touch-and-explore feature of LS_{palm}, which four participants noted as supporting learning, even for novices. For example:

> *"So someone who's new to it, I think they can pick it up fairly simple. […I like] the browsing part and being able to also open the apps. I can go back the center of my palm and still browse through that." (P5)*

Efficiency. We expected that *LI* would be perceived as less efficient than the two *LS* interfaces, since *LI* required a longer sequence of inputs (several swipes) compared to selecting a category by simply tapping its dedicated location. However, as seen in Figure 5, participants were roughly split between *LI* and LS_{palm} as the most efficient interaction technique, while only one participant selected LS_{body}. Seven participants commented that LS_{body}'s efficiency was impacted by the large motion required to point to different locations, which was especially of concern before learning the mappings. For instance:

> *"[With LS_{palm}] it's easy to search through. Because if it's not here, it's somewhere in this area. But [with LS_{body}] if you have it on your wrist and then your palm and on your ear, you gotta remember which one you need to touch." (P12)*

Social acceptability. Social acceptability is an important concern for wearable interaction [31] and for on-body interaction for VI users specifically, as seen both in prior work [25] and in Study 1. Nine participants felt they would be most comfortable using either *LI* or LS_{palm} in public compared to LS_{body} because the latter was less discreet. The remaining three participants felt that all three interaction techniques would be equally appropriate or inappropriate for public use. For example, P1 stated *"[LS_{body}] looks weird and a bit obnoxious. I mean, not pleasing for the onlooker. […] So I don't feel comfortable using it publicly."* The specific body locations included in the interface, however, would likely impact this perception (as found in [25]). For example:

> *"The [outer] wrist, I think is okay. I guess the inside of the wrist is okay, too. […] But ear, nose, thigh, all those things would be more uncomfortable." (P6)*

Interaction on-the-go and input availability. We found that participants appreciated having multiple options for input locations, especially for mobile contexts where one hand is busy. For this reason, six participants liked that *LI* did not restrict the input location, whereas microinteractions in the *LS* conditions could only be accessed from their dedicated input location, which may not always be available. For instance, P10 preferred locations other than the palm for flexibility, commenting, *"If you have something in your hand, it's like for me, if I'm holding my cane, or I'm standing somewhere, I want to check the time, I can just tap on my wrist. So it's easier, because my hand is not always going to be free."*

Customization. At the end of the study, we asked participants if they would want to customize the prototype to support different apps. A variety of preferences emerged, in terms of both interaction location and tasks to support. While seven participants wanted to use the palm for input, others suggested different body locations. P12, for example, wished to use specific locations for different tasks, suggesting some possibilities that had not been evaluated in the study, such as fingertips for making phone calls, ear lobe for answering phone calls, wrist for checking the time, thigh for notifications, and even lips for activating voice input. Furthermore, participants differed in the applications they wanted to support. For instance, P2 wanted to perform only phone call-related tasks, while P8 wanted a wider variety of tasks, such as opening apps like a calendar, a clock, and games, as well as checking notifications and heart rate. This diversity suggests that customization may be beneficial.

4.5.3 Summary

Our interview findings reinforced the survey responses regarding current support for microinteractions, specifically strategic use of voice input. The perceived trade-offs between three interaction techniques from the real-time system evaluation revealed VI users' needs and preferences for on-body interaction in terms of ease of use (*e.g.*, touch-and-explore to support learnability), efficiency (*e.g.*, downside of LS_{body}'s large movements), and use in public (*e.g.*, differences across body locations). The ability to support use on-the-go by supporting flexibility of input was also important.

5. Discussion

We reflect on the implications of our findings, focusing on how to better support microinteractions for people with visual impairments.

5.1 Accessible Microinteractions

Our findings suggest that efficient manual input for microinteractions is not well-supported for users with visual impairments. Only two microinteractions, checking the time and controlling a music player, were reported as taking < 5 seconds for most of our VI survey respondents. As seen in both the survey and interviews, voice input is a valuable tool for addressing the inefficiencies of screenreaders for microinteractions. This finding reflects previous more general studies showing that voice input is more popular for VI users than for sighted users [5,39]. However, because voice input may not always be appropriate due to background noise or privacy concerns [1,5,39], it is also important to develop more efficient manual input approaches to better support non-visual microinteractions.

Tasks that use mobile technology to increase the accessibility of the physical world, such as OCR, currency detection, and object recognition, arose rarely in the survey and interviews. This was, perhaps, because our focus was on tasks that are frequent *and* fast. Responses from three interview participants support this possibility: all three mentioned using their phone frequently for accessibility-specific tasks but not efficiently enough to qualify as a microinteraction; only one of the 61 survey respondents mentioned such a task (currency detection) as taking < 10 seconds. There is thus an opportunity for new wearable devices or even redesigned smartphone apps or operating systems to much more efficiently support accessibility-specific physical world tasks.

As already mentioned, wearable devices are particularly well-suited to support microinteractions [4,22,24]—whether smartwatches, on-body interaction, or other devices. However, differences in how VI and sighted users perceive smartwatches suggest that the ideal wearable form factor for each user group may differ, particularly in terms of the bulk and size needed to support visual versus auditory output modalities. VI users were more likely than sighted users to perceive the small size of a smartwatch to be an advantage, whereas sighted users, who rely on the visual output of the display, were more likely to see this size as a disadvantage. Still, some VI users were concerned about the touchscreen size from a gestural input perspective.

5.2 On-body Input for Microinteractions

While a variety of approaches—such as new wearable devices and redesigned touchscreen interactions—should be considered for future work on non-visual microinteractions, our findings provide specific guidance for implementing an on-body solution. Overall,

there were tradeoffs between the three on-body interaction techniques in efficiency, learnability, use on-the-go, and social acceptability. For efficiency, we had expected that the location-specific interaction techniques would fare well because they allowed for direct access to menu items or applications rather than requiring a series of swipes to find a specific item. However, participants felt that the two techniques that supported interaction within a small space were the most efficient: location-specific palm gestures and location-independent gestures, the latter of which participants often did on their hand. This finding suggests that quick access to all information within a constrained, discreet space may overshadow the efficiency advantages of direct selection at more distributed locations on the body, as well as the utility of semantic mappings (*e.g.*, wrist for time, ear for volume).

Learnability of the interface was also a concern. Half the participants mentioned that a strength of the location-independent gestures was not having to memorize the mappings between locations and menu items. Although we had expected that the touch-to-explore feature of location-specific gestures would allow users to discover items and support novice use, this was not seen in our study. However, a larger set of menu items and a longer-term study (rather than a single session) would likely affect perceptions of learnability. As users transition to experts, they may better appreciate the speed with which they can directly access items at specific hand and/or body locations. Thus, enabling both location-independent and location-specific gestures may best support both novice and expert users, similar to how iOS and Android support both gesture-based navigation as well as exploration by touch.

The three interaction techniques vary in their ability to support use on-the-go since not all input locations will always be available (*e.g.*, hand or thigh). Thus, in mobile contexts, input locations that move less when the whole body is in motion (*e.g.*, chest *vs.* leg) or that can be used even when one hand is busy (*e.g.*, holding a cane or bag) may be preferred. Alternatively, location-independent gestures, which are not restricted to a specific location, would allow for the flexibility to adapt the interaction to any given body posture or movement. As with issues of learnability, a combination of location-independent and location-specific input may also be useful to explore for use on-the-go. Regardless, complementing manual input with voice input is critical for efficient, on-the-go interaction.

Finally, social acceptability is a common concern with wearable assistive technology for VI users [30,39]. This concern arose with on-body interaction in Study 1 but was less pronounced in Study 2 where participants had experience using a real-time prototype. The gestures at locations across the body (*e.g.*, ear, thigh, wrist) were seen as less appropriate for use in public than the other two interaction techniques, which supported much smaller input areas. Socially acceptability concerns may be mitigated by supporting location-independent gestures as users can choose any input location that they consider discreet or by mapping all applications to a concentrated area that can support subtle gestures. Allowing users to customize body locations may also be useful, as appropriate locations may differ from one person to another (*e.g.*, due to gender or personal preference).

5.3 Limitations

A limitation of the online survey and first portion of the in-person study is that the data on microinteraction timing and frequency of use is based on subjective estimates. As such, it important for future work to complement these findings with log data of actual use. Another survey limitation is that the VI user group was older than the sighted user group, which may have impacted results. For the in-person study, while our system offered real-time interaction, it was still an early prototype and had drawbacks that could have impacted user perceptions. Most notably, the wearable components were bulkier than we envision a future system would be and the system required a lengthy training process (~30 minutes) to collect training images of the user's hands and body. As well, the system included only five application categories, which were accessed via location-specific taps or location-independent swipes and taps. A larger number of top-level items in the interface may change how users view the efficiency tradeoffs between the different interaction techniques. Finally, a longer-term evaluation of the on-body interaction techniques may yield different tradeoffs by addressing novelty effects and allowing users to develop expertise.

6. CONCLUSION

Our goal is to support not only *accessible* mobile interaction for users with visual impairments but *efficient* interaction. We investigated how well microinteractions are supported on smartphones and smartwatches for people with visual impairments compared to sighted users. Our findings suggest that manual input does not sufficiently support efficient non-visual microinteractions, highlighting the need for new interaction approaches. Speech input was a commonly used workaround for VI users in our online survey, but is not always appropriate due to privacy and noise concerns. As an alternative manual input approach, we evaluated three different on-body interaction techniques, which revealed tradeoffs in terms of perceived efficiency, learnability, social acceptability and ability to use on the go. Future work should examine to what extent it is possible to combine the strengths of these different on-body interaction techniques into a single system to ultimately allow for efficient and socially acceptable input.

7. ACKNOWLEDGEMENTS

We thank our participants, David Ross and the Atlanta VAMC, and the National Library Service for the Blind and Physically Handicapped. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

8. REFERENCES

1. Abdolrahmani, A., Kuber, R., and Hurst, A. 2016. An Empirical Investigation of the Situationally-Induced Impairments Experienced by Blind Mobile Device Users. *ACM W4A '16*: 21.
2. Adams, D., Morales, L., and Kurniawan, S. 2013. A qualitative study to support a blind photography mobile application. *ACM PETRA '13*, 1–8.
3. Ashbrook, D. L. 2010. Enabling Mobile Microinteractions. *PhD Thesis*.
4. Ashbrook, D. L., Clawson, J. R., Lyons, K., Starner, T. E., and Patel, N. 2008. Quickdraw: The Impact of Mobility and On-Body Placement on Device Access Time. *ACM CHI* '08, 219–222.
5. Azenkot, S. and Lee, N. B. 2013. Exploring the use of speech input by blind people on mobile devices. *ACM ASSETS '13*: 11.
6. Azenkot, S. Rector, K., Ladner, R., and Wobbrock, J. O. 2012. PassChords: Secure Multi-Touch Authentication for Blind People. *ACM ASSETS '12*, 159–166.
7. Böhmer, M., Hecht, B., Schöning, J., Krüger, A., and Bauer, G. 2011. Falling Asleep with Angry Birds, Facebook and Kindle: A Large Scale Study on Mobile Application Usage. *ACM MobileHCI '11*, 47–56.
8. Bonner, M. N., Brudvik, J. T., Abowd, G. D., and Edwards, W.

K. 2010. No-look Notes: Accessible Eyes-free Multi-touch Text Entry. *Springer-Verlag Pervasive '10*, 409–426.

9. Braun V. and Clarke, V. 2006. Using Thematic Analysis in Psychology. *Qualitative research in psychology* 3, 2: 77–101.
10. Chen, X. A., Marquardt, N., Tang, A., Boring, S., and Greenberg, S. 2012. Extending a mobile device's interaction space through body-centric interaction. *ACM MobieHCI '12*, 151–160.
11. Dakopoulos. D. and Bourbakis, N. G. 2010. 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.
12. Guerreiro, T., Gamboa, R., and Jorge, J. 2007. Mnemonical body shortcuts for interacting with mobile devices. *International Gesture Workshop*, 261–271.
13. Guerreiro, T., Lagoá, P., Nicolau, H., Gonçalves, D., and Jorge. J. 2008. From Tapping to Touching: Making Touch Screens Accessible to Blind Users. *IEEE MultiMedia '08*.
14. Gustafson, S., Rabe, B., and Baudisch, P. 2013. Understanding Palm-based Imaginary Interfaces: The Role of Visual and Tactile Cues when Browsing. *ACM CHI '13*, 889–898.
15. Gustafson S., Holz, C., and Baudisch, P. 2011. Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device. *ACM UIST '11*, 283–292.
16. Harrison, C., Ramamurthy, S., and Hudson, S. E. 2012. On-body Interaction: Armed and Dangerous. *ACM TEI '12*, 69–76.
17. Harrison, C., Tan, D., and Morris, D. 2010. Skinput: Appropriating the Body As an Input Surface. *ACM CHI '10*, 453–462.
18. Hoggan, E., Brewster S. A., and Johnston, J. 2008. Investigating the Effectiveness of Tactile Feedback for Mobile Touchscreens. *ACM CHI '08*, 1573–1582.
19. Hruschka, D. J., Schwartz, D., St.John D. C., Picone-Decaro, E., Jenkins, R. A., and Carey, J. W. 2004. Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research. *Field Methods* '04: 16(3), 307–331.
20. Kane, S. K., Bigham, J. P., and Wobbrock, J. O. 2008. Slide Rule: Making Mobile Touch Screens Accessible to Blind People Using Multi-touch Interaction Techniques. *ACM ASSETS '08*, 73–80.
21. Kane, S. K., Jayant, C., Wobbrock, J. O., and Ladner, R. E. 2009. Freedom to Roam: A Study of Mobile Device Adoption and Accessibility for People with Visual and Motor Disabilities. *ACM ASSETS '09*, 115–122.
22. Loclair, C., Gustafson, S., and Baudisch, P. 2010. PinchWatch: A Wearable Device for One-Handed Microinteractions. *ACM MobieHCI '10:*10
23. McGookin, D., Brewster, S., and Jiang, W. 2008. Investigating Touchscreen Accessibility for People with Visual Impairments. *ACM NordiCHI '08*, 298–307.
24. Nanayakkara, S., Shilkrot, R., Yeo, K. P., and Maes, P. 2013. EyeRing: A Finger-worn Input Device for Seamless Interactions with Our Surroundings. *ACM AH '13*, 13–20.
25. Oh, U. and Findlater, L. 2014. Design of and Subjective Responses to on-Body Input for People With Visual Impairments. *ACM ASSETS '14*: 8.
26. Oh, U. and Findlater, L. 2015. A Performance Comparison of On-hand versus On-phone Non-Visual Input by Blind and Sighted Users. *ACM TACCESS '15:* 7(4), 14.
27. OrCam Technologies Ltd. http://www.orcam.com. .
28. Oulasvirta, A., Tamminen, S., Roto, V., and Kuorelahti, J. 2005. Interaction in 4-Second Bursts: The Fragmented Nature of Attentional Resources in Mobile HCI. *ACM CHI '05*, 919–928.
29. Pizza, S., Brown, B., McMillan, D., and Lampinen, A. 2016. Smartwatch in Vivo. *ACM CHI '16*, 5456–5469.
30. Profita, H., Albaghli, R., Findlater, L., Jaeger, P., and Kane, S. K. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. *ACM CHI '16*, 4884–4895.
31. Profita, H. P., Clawson, J., Gilliland, S., Zeagler, C., Startner, T., Budd, J., and Do, E.Y. 2013. Don't Mind Me Touching My Wrist: A Case Study of Interacting with On-body Technology in Public. *ACM ISWC '13*, 89–96.
32. Shilkrot, R., Huber, J., Ee, W. M., Maes, P., and Nanayakkara, S. C. 2015. FingerReader: A Wearable Device to Explore Printed Text on the Go. *ACM CHI '15 Extended Abstracts*, 2363–2372.
33. Shinohara, K. and Wobbrock, J. O. 2011. In the Shadow of Misperception: Assistive Technology Use and Social Interactions. *ACM CHI '11*, 705–714.
34. Starner, T. E., Snoeck, C. M., Wong, B. A., and McGuire, R. M. 2004. Use of Mobile Appointment Scheduling Devices. *ACM CHI '04 Extended Abstracts*, 1501–1504.
35. Stearns, L., Du, R., Oh, U., Jou, C., Findlater, L., Ross, D. A., and Froehilich, J. E. 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind Persons in Reading Printed Text Using Finger-Mounted Cameras. *ACM TACCESS '16:*9(1), 1–38.
36. Stearns, L., Oh, U., Cheng, B. J., Findlater, L., Ross, D., Chellappa, R., and Froehlich, J.E. 2016. Localization of Skin Features on the Hand and Wrist from Small Image Patches. *IEEE ICPR '16*, 1003–1010.
37. WebAIM. *Screen Reader User Survey #4. http://webaim.org/projects/screenreadersurvey4/#demographics*.
38. Weigel, M., Mehta, V., and Steimle, J. 2014. More Than Touch: Understanding How People Use Skin As an Input Surface for Mobile Computing. *ACM CHI '14*, 179–188.
39. Ye H., Malu, M., Oh, U., and Findlater, L. 2014. Current and Future Mobile and Wearable Device Use by People with Visual Impairments. *ACM CHI '14*, 3123–3132.

A Multi-Modal Approach for Blind and Visually Impaired Developers to Edit Webpage Designs

Venkatesh Potluri, Liang He, Christine Chen, Jon E. Froehlich, Jennifer Mankoff
Paul G. Allen School of Computer Science & Engineering, University of Washington
{vpotluri, lianghe, chenc55, jonf, jmankoff}@cs.washington.edu

ABSTRACT
Blind and visually impaired (BVI) individuals are increasingly creating visual content online; however, there is a lack of tools that allow these individuals to modify the visual attributes of the content and verify the validity of those modifications. In this poster paper, we discuss the design and preliminary exploration of a multi-modal and accessible approach for BVI developers to edit visual layouts of webpages while maintaining visual aesthetics.

Author Keywords
Accessible programming; visual design; web; blind and visually impaired.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces

INTRODUCTION
Content creators who are blind and visually impaired (BVI) are actively building interfaces that contain visual elements, such as blogs and personal websites [6]. While these interfaces are expected to have good visual design, the necessary tools and information to build visually pleasing [3,12] layouts are inaccessible to BVI developers [11]. To address this problem, we propose a multi-modal approach that allows BVI developers to edit webpages without breaking visual aesthetics.

Prior research in accessible programming tools for BVI developers has addressed challenges such as effective code navigation for BVI developers [2] and proposed new integrated development environment (IDE) techniques to make programming more accessible [4,16,18]. These approaches, however, do not address the accessibility barriers specific to editing the visual attributes of user interfaces (UIs). To bridge the gap, *Borka* [5], an addon for the NVDA screen reader [14], informs the developer of the location, height, and width of web UI elements in pixels. Li *et al.* [11] built a self-voicing application on a tablet and tactile graphics to support visual layout editing. While related, our approach is unique in that it allows for web design edits both through an IDE as well as directly on the rendered output of a touchscreen—edits made via either mode are automatically assessed via a design validator.

Specifically, we introduce a multi-modal UI modification approach that combines an *accessible code editor* along with a *touchscreen* and *controller* to directly access, edit, and validate UI designs (Figure 1.). With our system, BVI developers edit CSS code from an existing template either using an accessible IDE or through a touchscreen. On the touchscreen, the BVI developer can *directly explore* the design via touch and use touch gestures to make *direct edits.* When an edit is made, it is assessed by the validation engine for visual design adherence. The screen reader that runs alongside the IDE announces the outcome of the edit. For example, if visual design guidelines are violated, the changes are reverted on both the IDE and the touchscreen, and the screen reader announces the design guideline violations.

In a preliminary pilot study, we found that, with minimal training, a BVI developer could effectively make meaningful visual edits with real-time validation using our system.

MULTI-MODAL SYSTEM
To explore our multi-modal approach, we created a three-part system (Fig. 1): an *accessible canvas* that allows BVI developers to use touch and gestures to modify the visual properties of web UI elements, a *code editor* that supports direct code modifications, and a *controller* that processes proposed updates from either the canvas or code editor and then checks if the updates violate design guidelines. Please refer to our supplementary video for a demonstration.

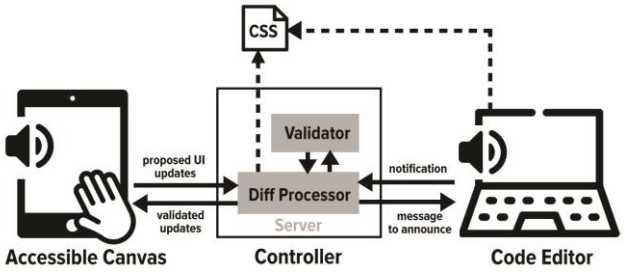


Figure 1. Multi-modal system overview. See video demo.

Design Guidelines
While visual design is subjective and hard to quantify [3], we ensure that the edits BVI developers make with our system adhere to an initial set of design guidelines extending from web design standards [1,12]:

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ASSETS '19, October 28–30, 2019, Pittsburgh, PA, USA
© 2019 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-6676-2/19/10.
https://doi.org/10.1145/3308561.3354626

- *Spacing consistency*. The horizontal and vertical spacing between elements at the same level should be consistent.
- *Typeface consistency*. Significant variation in fonts can be distracting and confusing. A maximum of three different font types on a page is enforced.
- *Color consistency*. UI elements of the same type should share the same color.

Accessible Canvas

To support exploration and manipulation of the web UI elements, our system displays a representation of the webpage (HTML and CSS) on a touchscreen (an iPad Air running iOS 12.2). The user can navigate and edit their design directly using gestures (Table 1), which are implemented via Apple UIkit Gestures. Since performing custom gestures in iOS conflicts with built-in VoiceOver, we built the accessible canvas to be self-voicing. The canvas announces the element the user touches and provides audio feedback on the user edits, which are performed by gestures (Table 1). Our gestures are based on prior work, such as risk-free exploration suggested by Kane *et al.* [7] and leveraging the familiarity of edge elements [8].

Operation	Touch and Gestures
Navigating	Touch and move
Aligning to one side	One-finger double taps and two-finger swipe
Style change mode	**Two-finger triple taps**
Changing width	Horizontal edge tap and one-finger moving horizontally
Changing height	Vertical edge tap and one-finger moving vertically
Margin/padding	**Four-finger long press**
Changing margin/padding	One-finger swipe to one side and one-finger moving vertically or horizontally
Background	**Three-finger long press**
Changing color	One-finger swipe left or right
Foreground	**Two-finger long press**
Changing font color	One-finger swipe left or right
Changing font size	One-finger swipe up and down
Changing font type	Two-finger swipe left or right
Submitting changes	**Three-finger triple taps**

Table 1. The full list of gestures used for navigation, alignment, and adjustment of element style.

Code Editor

In addition to making touchscreen edits, the developer can also directly edit the webpage CSS in Microsoft's Visual Studio Code (VS Code) using a keyboard and a screen reader. When the developer saves changes to the CSS file in the editor, a custom VS Code extension notifies the controller about the new code edits. Additionally, the extension receives messages from the controller (on the validity of the edits made in the editor or the accessible canvas). These messages are announced by the screen reader running alongside the editor.

Controller

The controller has two parts: a *diff processor* and *validator*. The *diff processor* receives proposed changes from the accessible canvas or a notification from the code editor, updates a working copy of the original CSS file on the controller, and then evaluates the updates with the *validator*. If the proposed change passes validation, the working copy is checkpointed, the accessible canvas is updated, and the code editor receives an accept message (announced by the screen reader). If the change violates a design guideline, the processor reverts the CSS to the last checkpoint, leaving the accessible canvas unchanged. A rejection message, including the violated design guideline, is then reported in the code editor (see the video).

To check for adherence to design guidelines, the validator maintains an internal representation of the visual attributes of the webpage elements (from the CSS file) and the hierarchy of elements (from the HTML). For spacing consistency, we used the margin attributes of adjacent elements for validation. For example, the margins of horizontally distributed elements should be the same. We examined the font and color attributes for typeface consistency and color consistency, respectively.

PRELIMINARY USER STUDY

We conducted a preliminary user study with a congenitally blind professional software engineer (age 24; male). The study had three parts: a semi-structured interview on current webpage development strategies, an evaluation of our system through three tasks (*i.e.*, changing the typeface of a paragraph, increasing the bottom margin for a list of hyperlinks, and aligning an image and a paragraph), and a post-study interview regarding usability. The participant completed all three tasks successfully and agreed that the gestures are usable, although he had trouble memorizing gestures for operations. He could understand the outcomes of the tasks—edits either passed or failed design guideline violation check. The participant expressed the need to have more control, for example, he wanted to have access to the history of changes that lead to a rejection.

CONCLUSION AND DISCUSSION

We presented an initial exploration of enabling BVI developers to edit webpage design using a multi-modal approach. Our system allows BVI developers to change visual aspects of web UI elements either by performing gestures on a touchscreen or by directly modifying code in an editor. We plan to enhance our validator by integrating the design guidelines with VizAssert [15], a framework that verifies accessibility of webpages. We are also interested in exploring machine learning techniques both for validation as well as to suggest alternative designs (*e.g.*, based on design trends suggested in [10]) [17]. Finally, we will investigate the discoverability and learnability of our current gesture set through a formal study. Besides gestures, we will assess viability of accessible interaction techniques in [9] and alternative input devices (*e.g.*, touchpads) for modifying visual attributes of interfaces.

ACKNOWLEDGEMENTS

This work is partly supported by NSF Award IIS-1836813 and the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617. We thank Michael Ernst and Pavel Panchekha for their feedback on the development of our system.

REFERENCES

1. Maneesh Agrawala, Wilmot Li, and Floraine Berthouzoz. 2011. Design principles for visual communication. *Commun. ACM* 54, 4 (April 2011), 60-69. DOI: https://doi.org/10.1145/1924421.1924439
2. Khaled Albusays, Stephanie Ludi, and Matt Huenerfauth. 2017. Interviews and Observation of Blind Software Developers at Work to Understand Code Navigation Challenges. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '17). ACM, New York, NY, USA, 91-100. DOI: https://doi.org/10.1145/3132525.3132550
3. Ahamed Altaboli and Yingzi Lin. 2011. Objective and subjective measures of visual aesthetics of website interface design: the two sides of the coin. In *International Conference on Human-Computer Interaction*, pp. 35-44. Springer, Berlin, Heidelberg
4. Catherine M. Baker, Lauren R. Milne, and Richard E. Ladner. 2015. StructJumper: A Tool to Help Blind Programmers Navigate and Understand the Structure of Code. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 3043-3052. DOI: https://doi.org/10.1145/2702123.2702589
5. Andy Borka. 2019. NVDA Community Addons / Addons / Developer Toolkit. Retrieved July 8, 2019 from https://addons.nvda-project.org/addons/developerToolkit.uk.html
6. Dolphin. 2016. Blogging when you're blind or visually impaired. Retrieved July 9, 2019 from https://yourdolphin.com/news?id=223
7. Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility* (Assets '08). ACM, New York, NY, USA, 73-80. DOI: https://doi.org/10.1145/1414471.1414487
8. Shaun K. Kane, Jacob O. Wobbrock, and Richard E. Ladner. 2011. Usable gestures for blind people: understanding preference and performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 413-422. DOI: https://doi.org/10.1145/1978942.1979001
9. Rushil Khurana, Duncan McIsaac, Elliot Lockerman, and Jennifer Mankoff. 2018. Nonvisual Interaction Techniques at the Keyboard Surface. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). ACM, New York, NY, USA, Paper 11, 12 pages. DOI: https://doi.org/10.1145/3173574.3173585
10. Ranjitha Kumar, Arvind Satyanarayan, Cesar Torres, Maxine Lim, Salman Ahmad, Scott R. Klemmer, and Jerry O. Talton. 2013. Webzeitgeist: design mining the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13). ACM, New York, NY, USA, 3083-3092. DOI: https://doi.org/10.1145/2470654.2466420
11. Jingyi Li, Son Kim, Joshua A. Miele, Maneesh Agrawala, and Sean Follmer. 2019. Editing Spatial Layouts through Tactile Templates for People with Visual Impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). ACM, New York, NY, USA, Paper 206, 11 pages. DOI: https://doi.org/10.1145/3290605.3300436
12. Laura Martin. 2017. Typography Elements Everyone Needs to Understand. Retrieved July 8, 2019 from https://medium.com/gravitdesigner/typography-elements-everyone-needs-to-understand-5fdea82f470d
13. Microsoft. Visual Studio Code. Retrieved July 9, 2019 from https://code.visualstudio.com/
14. NV Access. Non Visual Desktop Access. Retrieved July 9, 2019 from https://www.nvaccess.org/about-nv-access/
15. Pavel Panchekha, Adam T. Geller, Michael D. Ernst, Zachary Tatlock, and Shoaib Kamil. 2018. Verifying that web pages have accessible layout. In *Proceedings of the 39th ACM SIGPLAN Conference on Programming Language Design and Implementation* (PLDI 2018). ACM, New York, NY, USA, 1-14. DOI: https://doi.org/10.1145/3192366.3192407
16. Venkatesh Potluri, Priyan Vaithilingam, Suresh Iyengar, Y. Vidya, Manohar Swaminathan, and Gopal Srinivasa. 2018. CodeTalk: Improving Programming Environment Accessibility for Visually Impaired Developers. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). ACM, New York, NY, USA, Paper 618, 11 pages. DOI: https://doi.org/10.1145/3173574.3174192
17. Venkatesh Potluri, Tadashi Grindeland, Jon E. Froehlich, Jennifer Mankoff (2019). AI-Assisted UI Design for Blind and Low-Vision Creators. In the *ASSETS'19 Workshop: AI Fairness for People with Disabilities.*
18. Emmanuel Schanzer, Sina Bahram, and Shriram Krishnamurthi. 2019. Accessible AST-Based Programming for Visually-Impaired Programmers. In *Proceedings of the 50th ACM Technical Symposium on*

Computer Science Education (SIGCSE '19). ACM, New York, NY, USA, 773-779. DOI: https://doi.org/10.1145/3287324.3287499

The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind

Lee Stearns[1], Ruofei Du[1], Uran Oh[1], Yumeng Wang[2], Leah Findlater[3], Rama Chellappa[4], Jon E. Froehlich[1]

[1]Computer Science, [2]School of Architecture,
[3]College of Information Studies, [4]Electrical and Computer Engineering
University of Maryland, College Park
{lstearns, ruofei, uranoh, yumeng, leahkf, chella, jonf}@umd.edu

Abstract. We introduce the preliminary design of a novel vision-augmented touch system called *HandSight* intended to support activities of daily living (ADLs) by sensing and feeding back non-tactile information about the physical world *as it is touched*. Though we are interested in supporting a range of ADL applications, here we focus specifically on reading printed text. We discuss our vision for HandSight, describe its current implementation and results from an initial performance analysis of finger-based text scanning. We then present a user study with four visually impaired participants (three blind) exploring how to continuously guide a user's finger across text using three feedback conditions (haptic, audio, and both). Though preliminary, our results show that participants valued the ability to access printed material, and that, in contrast to previous findings, audio finger guidance may result in the best reading performance.

Keywords: Accessibility, Wearables, Real-time OCR, Text Reading for Blind

1 Introduction

Over 285 million people have visual impairments (VI) worldwide—including 39 million who are blind—that can affect their ability to perform activities of daily living [1]. In many countries, such as the US [2], VI prevalence is increasing due to aging populations. While previous research has explored combining mobile cameras and computer vision to support people with VI for at-a-distance information tasks such as navigation (*e.g.,* [3]–[8]), facial recognition (*e.g.,* [9]–[12]), and spatial perception (*e.g.,* [13]–[15]), they fail to support proximal information accessed through touch.

We are pursuing a new approach: a vision-augmented touch system called *HandSight* that supports activities of daily living (ADLs) by sensing and feeding back non-tactile information about the physical world *as it is touched*. Although still at an early stage, our envisioned system will consist of tiny CMOS cameras ($1\times1mm^2$) and micro-haptic actuators mounted on one or more fingers, computer vision and machine learn-

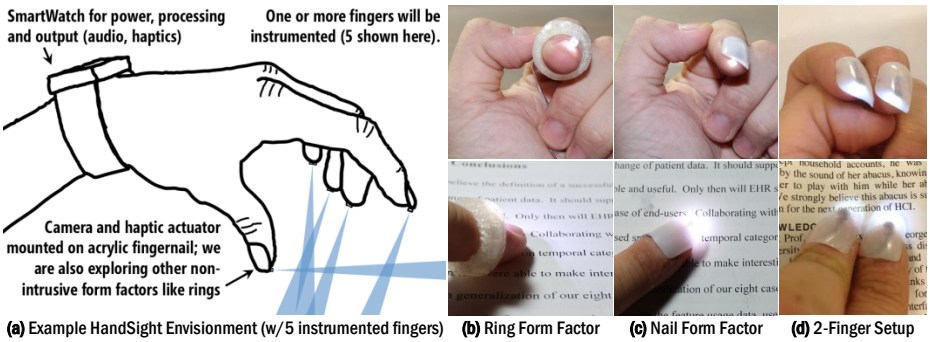


Figure 1: HandSight uses a 1×1mm² *AWAIBA NanEye 2C* camera developed for minimally invasive surgeries (*e.g.*, endoscopies) that can capture 250×250px images at 44fps. The above images: (a) a design mockup and (b-d) early form factors with the NanEye camera. In this paper, we explore a single camera implementation with a ring form factor in a text reading context (see Figure 2).

ing algorithms to support fingertip-based sensing, and a smartwatch for processing, power, and speech output; see Figure 1. Since touch is a primary and highly attuned means of acquiring information for people with VI [16], [17], we hypothesize that collocating the camera with the touch itself will enable new and intuitive assistive applications.

While we are interested in supporting a range of applications from object recognition to color identification, in this paper we focus on the challenge of reading printed text. Our overarching goal is to allow blind users to touch printed text and receive speech output in real-time. The user's finger is guided along each line via haptic and non-verbal audio cues. At this stage, our research questions are largely exploratory, spanning both human-computer interaction (HCI) and computer vision: How can we effectively guide the user's finger through haptic and auditory feedback to appropriately scan the target text and notify them of certain events (*e.g.*, start/end of line or paragraph reached)? How accurately can optical character recognition (OCR) be achieved at a speed that is responsive to the user's touch? How do the position, angle, and lighting of the finger-mounted camera affect OCR performance?

To begin examining these questions, we pursued two parallel approaches. For the computer vision questions, we developed an early HandSight prototype along with efficient algorithms for perspective and rotation correction, text detection and tracking, and OCR. We present preliminary evaluations and demonstrate the feasibility of our envisioned system. For the HCI-related questions, we developed a custom test apparatus on the Apple iPad that simulates the experience of using HandSight, but provides us with additional experimental control and allows us to more precisely track the user's finger in response to feedback conditions. Using this setup, we report on a preliminary evaluation with four VI participants (three blind) across three finger guidance conditions: *audio*, *haptics*, and *audio+haptics*. Findings suggest that audio may be the most intuitive feedback mechanism of the three.

Compared to the majority of emerging computer vision systems to support VI users, which use head- or chest-mounted cameras (*e.g.,* [4], [12], [18]), our system offers two primary advantages: (i) *collocation of touch, sensing, and feedback*, potentially enabling more intuitive interaction and taking advantage of a VI individual's high tactile acuity [16], [17]; (ii) *unobtrusive, always-available interaction* that allows for seamless switching between the physical world and vision-augmented applications. As a finger-mounted approach, our work is most similar to [19]–[21], described next.

2 Related Work

Scientists have long sought to support blind people in reading printed text (for a review, see [22], [23]). Many early so-called "reading machines for the blind" used a sensory substitution approach where the visual signals of words were converted to non-verbal auditory or tactile modalities, which were complicated to learn but accessible. Two such examples include the *Optophone*, which used musical chords or 'motifs' [24] and the *Optacon*, which used a vibro-tactile signal [25], [26]. With advances in sensing, computation, and OCR, modern approaches attempt to scan, recognize, and read aloud text in real-time. This transition to OCR and speech synthesis occurred first with specialized devices (*e.g.,* [27]–[29]), then mobile phones (*e.g.,* [30], [31]), and now wearables (*e.g.,* [12], [21]). While decades of OCR work exist (*e.g.,* [32]–[35]), even state-of-the-art reading systems become unusable in poor lighting, require careful camera framing [36], [37], and do not support complex documents and spatial data [38]. Because HandSight is self-illuminating and co-located with the user's touch, we expect that many of these problems can be mitigated or even eliminated.

As a wearable solution, HandSight is most related to *OrCam* [12] and *FingerReader* [21]. OrCam is a commercial head-mounted camera system designed to recognize objects and read printed text in real-time (currently in private beta testing). Text-to-speech is activated by a pointing gesture in the camera's field-of-view. While live demonstrations with sighted users have been impressive (*e.g.,* [39], [40]), there is no academic work examining its effectiveness with VI users for reading tasks. The primary distinctions between HandSight and OrCam are, first, hand-mounted versus head-mounted sensing, which could impact camera framing issues and overall user experience. Second, HandSight supports direct-touch scanning compared to OrCam's indirect approach, potentially allowing for increased control over what is read and reading speed as well as increased spatial understanding of a page/object. Regardless, the two approaches are complementary, and we plan to explore a hybrid in the future.

More closely related to HandSight, FingerReader [21] is a custom finger-mounted device with vibration motors designed to read printed text by direct line-by-line scanning with the finger. Reported evaluations [21] of FingerReader are limited to a very small OCR assessment under unspecified "optimal" conditions, and a qualitative user

study with four blind participants. The participants preferred haptic to audio-based finger guidance; this finding is the opposite of our own preliminary results, perhaps due to differences in how the audio was implemented (theirs is not clearly described). Further, our study extends [21] in that we also present user performance results.

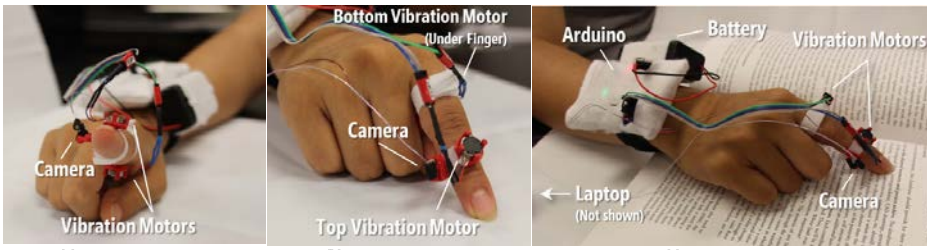
In terms of finger guidance, haptic and audio feedback have been used in numerous projects to guide VI users in exploring non-tactile information or tracing shapes. Crossan and Brewster [41], for example, combined pitch and stereo sonification with a force feedback controller to drag the user along a trajectory, and found that performance was higher with audio and haptic feedback than haptic feedback alone. Other approaches have included sonification and force feedback to teach handwriting to blind children [42], speech-based icons or "spearcons" [43], vowel sounds to convey radial direction [44], and use of primarily tactile feedback to transmit directional and shape data [45]–[47]. Our choice to vary pitch for audio-based line tracing feedback with HandSight is based on previous findings [41], [43], [48]. Oh *et al.* [48], *e.g.*, used sonification to support non-visual learning of touchscreen gestures; among a set of sound parameters tested (pitch, stereo, timbre, etc.), pitch was the most salient.

3 System Design

HandSight is comprised of three core components: sensors, feedback mechanisms, and a computing device for processing. Our current, early prototype is shown in Figure 2. Before describing each component in more detail, we enumerate six design goals.

3.1 Design Goals

We developed the following design goals based on prior work and our own experiences developing assistive technology: (1) *Touch-based rather than distal interaction.* Although future extensions to HandSight could examine distal interaction, our focus is on digitally augmenting the sense of touch. (2) *Should not hinder normal tactile function.* Fingers are complex tactile sensors [49], [50] that are particularly attuned in people with visual impairments [16], [17]; HandSight should not impede normal tactile sensation or hand function. (3) *Easy-to-learn/use.* Many sensory aids fail due to their complexity and high training requirements [22]; to ensure HandSight is approachable and easy to use, we employ an iterative, human-centered design approach. (4) *Always-available.* HandSight should allow for seamless transitions between its use and real-world tasks. There is limited prior work on so-called always-available input [20], [51]–[53] for blind or low-vision users. (5) *Comfortable & robust.* HandSight's physical design should support, not encumber, everyday activities. It should be easily removable, and water and impact resistant. (6) *Responsive & accurate.* HandSight



(a) Close-up front view (b) Close-up side view (c) Full system view

Figure 2: The current HandSight prototype with a NanEye ring camera, two vibration motors, and an Arduino. Finger rings and mounts are constructed from custom 3D-printed designs and fabric. Processing is performed in real-time on a laptop (not shown).

should allow the user to explore the target objects (*e.g.*, utility bills, books) quickly—the computer vision and OCR algorithms should work accurately and in real-time.

3.2 Hardware

Sensing Hardware. Our current prototype uses a single 1×1mm^2 *AWAIBA NanEye 2C* camera [54] that can capture 250×250 resolution images at 44 frames per second (fps). The NanEye was originally developed for minimally invasive surgical procedures such as endoscopies and laparoscopies and is thus robust, lightweight, and precise. The camera also has four LEDs coincident with the lens (2.5mm ring), which enables dynamic illumination control. The small size allows for a variety of finger-based form factors including small rings or acrylic nail attachments. In our current prototype, the camera is attached to an adjustable velcro ring via a custom 3D-printed clip.

Processing. For processing, we use a wrist-mounted Arduino Pro Micro with an attached Bluetooth module that controls the haptic feedback cues. The video feed from the camera is currently processed in real time on a laptop computer (our experiments used a Lenovo Thinkpad X201 with an Intel Core i5 processor running a single computation thread at approximately 30fps). Later versions will use a smartwatch (*e.g.*, Samsung Galaxy Gear [55]) for power and processing.

Feedback. HandSight provides continuous finger-guidance feedback via vibration motors, pitch-controlled audio, or both. Our current implementation includes two vibration motors that are **8mm diameter disks and 3.4mm thick** (Figure 2), though we are actively looking at other solutions (see Discussion). A text-to-speech system is used to read each word as the user's finger passes over it, and distinctive audio and/or haptic cues can be used to signal other events, such as end of line, start of line, etc.

3.3 CV Algorithm Design and Evaluation

Our current HandSight implementation involves a series of frame-level processing stages followed by multi-frame merging once the complete word has been observed.

Below, we describe our five stage OCR process and some preliminary experiments evaluating performance.

Stage 1: Preprocessing. We acquire grayscale video frames at ~40fps and 250x250px resolution from the NanEye camera (Figure 3). With each video frame, we apply four preprocessing algorithms: first, to correct radial and (slight) tangential distortion, we use standard camera calibration algorithms [56]. Second, to control lighting for the next frame, we optimize LED intensity using average pixel brightness and contrast. Third, to reduce noise, perform binarization necessary for OCR, and adapt to uneven lighting from the LED, we filter the frame using an adaptive threshold in a Gaussian window; finally, to reduce false positives, we perform a connected component analysis and remove components with areas too small or aspect ratios too narrow to be characters.

Stage 2: Perspective and Rotation Correction. The finger-based camera is seldom aligned perfectly with the printed text (*e.g.,* top-down, orthogonal to text). We have observed that even small amounts of perspective distortion and rotation can reduce the accuracy of our text detection and OCR algorithms. To correct perspective and rotation effects, we apply an efficient approach detailed in [56]–[58], which relies on the parallel line structure of text for rectification. We briefly describe this approach below.

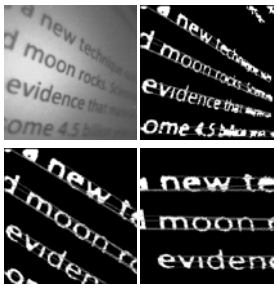


Figure 3: A demonstration of our perspective and rotation correction algorithm.

To identify potential text baselines, we apply a Canny filter that highlights character edges and a randomized Hough transform that fits lines to the remaining pixels. From this, we have a noisy set of candidate baselines. Unlikely candidates are filtered (*e.g.,* vertical lines, intersections that imply severe distortion). The remaining baselines are enumerated in pairs; each pair implies a potential rectification, which is tested on the other baselines. The baseline pair that results in the lowest line angle variance is selected and the resulting rectification is applied to the complete image.

More precisely, the intersection of each pair of baselines implies a horizontal vanishing point $V_x = l_1 \times l_2$ in homogeneous coordinates. If we assume the ideal vertical vanishing point $V_y = [0, 1, 0]^T$, then we can calculate the homography, H, that will make those lines parallel. Let $l_\infty = V_x \times V_y = [a, b, c]^T$ and calculate the perspective homography, H_p, using those values. The perspective homography makes the lines parallel, but does not align them with the x-axis. We must rotate the lines by an angle θ using a second matrix, H_r. The complete rectifying homography matrix becomes:

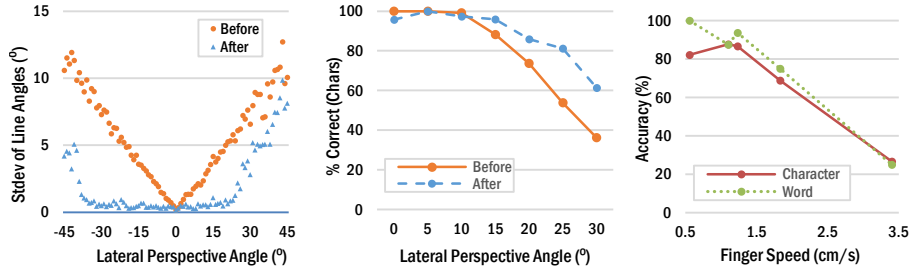
$$H = H_r H_p = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a/c & b/c & 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ a/c & b/c & 1 \end{bmatrix} \quad \textbf{(1)}$$

To investigate the effect of lateral perspective angle on performance, we performed a synthetic experiment that varied the lateral angle from -45° to 45° across five randomly selected document image patches. The raw rectification performance is shown in Figure 4a and the effect of rectification on character-level OCR is shown in Figure 4b accuracy (the algorithm for OCR is described below).

Stage 3: Text Detection. The goal of the text detection stage is to build up a hierarchy of text lines, words, and characters. This task is simplified because we assume the perspective and rotation correction in Stage 2 has made the text parallel to the *x*-axis. First, we split the image into lines of text by counting the number of text pixels in each row and searching for large gaps. Next, we split each line into words using an identical process on the columns of pixels. Gaps larger than 25% of the line height are classified as spaces between words. Finally, we segment each word into individual characters by searching for local minima in the number of text pixels within each column.

Stage 4: Character Classification. Real-time performance is important for responsive feedback, which prevents us from using established OCR engines such as Tesseract. Thus, we compute efficient character features (from [59]), and perform classification using a support vector machine (SVM). Each character candidate is centered and scaled to fit within a 32x32 pixel window, preserving the aspect ratio. The window is split into four horizontal and vertical strips, which are summed along the short axis to generate eight vectors of length 32 each. These vectors, along with the aspect ratio, perimeter, area, and thinness ratio make up the complete feature vector. The thinness ratio is defined as $T=4\pi(A/P^2)$ where A is the area and P is the perimeter. We compensate for the classifier's relatively low accuracy by identifying the top k most likely matches. By aggregating the results over multiple frames, we are able to boost performance.

Stage 5: Tracking and final OCR result output. The camera's limited field of view means that a complete word is seldom fully within a given frame. We must track the characters between frames and wait for the end of the word to become visible before we can confidently identify it. Character tracking uses sparse low-level features for efficiency. First, we extract FAST corners [60], and apply a KLT tracker [61] at their locations. We estimate the homography relating the matched corners using the random sample consensus [62]. After determining the motion between frames, we relate the lines, words, and individual characters by projecting their locations in the previous frame to the current frame using the computed homographies. The bounding boxes with the greatest amount of overlap after projection determine the matches. When the



(a) Performance of lateral perspective and rotation rectification algorithm.

(b) Effect of lateral perspective angle on accuracy (before and after correction).

(c) Effect of finger speed on character- and word-level accuracy.

Figure 4: Results from preliminary evaluations of our (a-b) Stage 2 algorithms and (c) the effect of finger speed on overall character- and word-level accuracy.

end of a word is visible, we sort the aggregated character classifications and accept the most frequent classification. This process can be improved by incorporating a language dictionary model, albeit at the expense of efficiency. A text-to-speech engine reads back the identified word.

To investigate the effect of finger movement speed on OCR accuracy, we recorded five different speeds using a single line of text. The results are presented in Figure 4c. With greater speed, motion blur is introduced, and feature tracking becomes less accurate. In our experience, a "natural" finger speed movement for sighted readers is roughly 2-3cm/s. So, with the current prototype, one must move slower than natural for good performance. We plan on compensating for this effect in the future using image stabilization and motion blur removal, as well as incorporating a higher frame rate camera (100fps).

4 User Study to Assess Audio and Haptic Feedback

Our current prototype implementation supports haptic and audio feedback, but how best to implement this feedback for efficient direct-touch reading is an open question. Ultimately, we plan to conduct a holistic user evaluation of the system to assess the combined real-time OCR and finger guidance for a variety of reading tasks. At this stage, however, our goal was to refine the finger guidance component of the system by conducting a preliminary evaluation of three types of feedback: (1) audio only, (2) haptic only, and (3) a combined audio and haptic approach. We conducted a user study with four visually impaired participants to collect subjective and performance data on these three types of feedback. To isolate the finger guidance from the current OCR approach, we used a custom iPad app that simulates the experience of using the full system.

4.1 Method

Participants. We recruited four VI participants; details are shown in Table 1. All four participants had braille experience, and three reported regular use of screen readers.

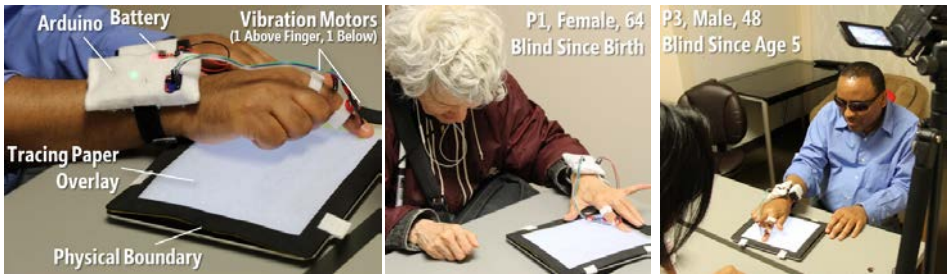
Test apparatus. The setup simulated the experience of reading a printed sheet of paper with HandSight (Figure 5). It consisted of the hand-mounted haptic component of the HandSight system controlled by an Arduino Micro, which was in turn connected via Bluetooth to an Apple iPad running a custom experimental app. The iPad was outfitted with a thin foam rectangle as a physical boundary around the edge of the screen to simulate the edge of a sheet of paper, and was further covered by a piece of tracing paper to provide the feel of real paper and to reduce friction. The app displayed text documents, guiding the user to trace each line of the document from left to right and top to bottom. As the user traced their finger on the screen, text-to-speech audio was generated, along with the following feedback guidance cues: start and end of a line of text, end of a paragraph, and vertical guidance for when the finger strayed above or below the current line. Lines were 36 pixels in height and vertical guidance began when the finger was more than 8 pixels above or below the vertical line center.

Feedback conditions tested. We compared three finger guidance options:

- *Audio only.* All guidance cues were provided through non-speech audio. The start and end of line cues were each a pair of tonal percussive (xylophone) notes played in ascending or descending order, respectively. The end of paragraph sound was a soft vibrophone note. When the user's finger drifted below or above a line, a continuous audio tone would be played to indicate that proper corrective movement. A lower tone (300 Hz) was played to indicate downward corrective movement (*i.e.,* the user was above the line). The pitch decreased at a rate of 0.83Hz/pixel to a minimum of 200Hz at 127 pixels above the line. A higher tone (500 Hz) was used to indicate upward corrective movement (up to a maximum of 600Hz with the same step value as before).
- *Haptic only.* The haptic feedback consists of two finger-mounted haptic motors, one on top and one underneath the index finger (see Section 3.2). Based on piloting within the research team, the motors were placed on separate finger segments (phalanges) so that the signal from each was easily distinguishable. To cue the start of a line, two short pulses played on both motors, with the second pulse more intense than the first; the reverse pattern indicated the end of a line. For the end of a paragraph, each motor vibrated one at a time, which repeated for a total of four pulses. For vertical guidance, when the finger strayed too high, the motor underneath the finger vibrated, with the vibration increasing in intensity from a

ID	Age	Gender	Handedness	Level of Vision	Years of Vision Loss	Diagnosed Med. Condition	Hearing Difficulties
P1	64	Female	Left	Totally blind	Since birth	Retinopathy of prematurity	N/A
P2	61	Female	Left	Totally blind	Since birth	Retinopathy of prematurity	Slight hearing loss
P3	48	Male	Right	Totally blind	Since age 5	N/A	N/A
P4	43	Female	Right	No vision one eye, 20/400 other eye	30 years	Glaucoma	N/A

Table 1: Background of the four user study participants.



(a) iPad test apparatus (b) Participant 1 (c) Participant 3

Figure 5: User study setup and test apparatus: (a) overview; (b-c) in use by two participants..

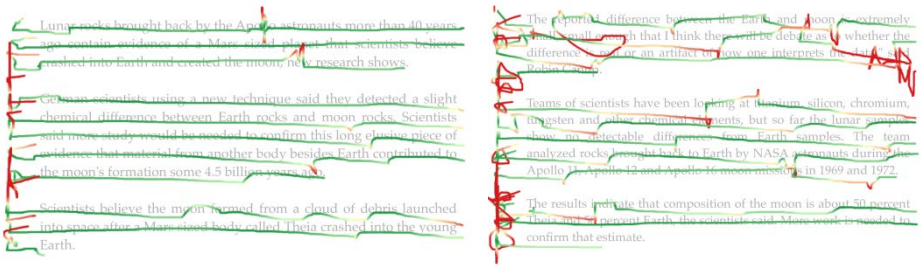
low perceivable value to maximum intensity, reached at 127 pixels above the line; below the line, the top motor vibrated instead (again with the maximum intensity reached at 127 pixels).

- *Combined audio-haptic.* This combined condition included all of the audio *and* haptic cues described above, allowing the two types of feedback to complement each other in case one was more salient for certain cues than the other.

Procedure. The procedure took up to 90 minutes. For each feedback condition, the process was as follows. First, we demonstrated the feedback cues for the start/end of each line, end of paragraph, and vertical guidance. Next, we loaded a training article and guided the user through the first few lines. Participants then finished reading the training article at their own pace. Finally, a test article was loaded and participants were asked to read through the text as quickly and accurately as possible. While we provided manual guidance as necessary to help participants read the *training* article (*e.g.*, adjusting their finger), no manual guidance was given during the *test* task. Four articles of approximately equivalent complexity were selected from *Voice of America* (a news organization), one for the training task and one to test each feedback condition; all articles had three paragraphs and on average 11.0 lines (*SD*=1.0) and 107.0 words (*SD*=13.5). The order of presentation for the feedback conditions was randomized per participant, while the test articles were always shown in the same order. Questions on ease of use were asked after each condition and at the end of the study. Sessions were video recorded, and all touch events were logged.

4.2 Findings

We analyzed subjective responses to the feedback conditions, and user performance based on logged touch events. Figure 6 shows a sample visualization from one participant (P1) completing the reading task in the *audio-only* and *haptic-only* conditions. Due to the small sample size, all findings in this section should be considered preliminary, but point to the potential impacts of HandSight and tradeoffs of different feedback.



(a) Participant 1 finger trace (audio only condition)

(b) Participant 1 finger trace (haptic only condition)

Figure 6: Our iPad test apparatus allowed us to precisely track and measure finger movement. Example trace graphs for Participant 1 (P1) across the audio- and haptic-only conditions are shown above (green is on line; red indicates off-line and guidance provided). These traces were also used to calculate a range of performance measures. For example, the average overall time to read a line for P1 took 11.3s (*SD*=3.9s) in the audio condition and 18.9s (*SD*=8.3s) in the haptic condition. The average time to find the beginning of the next line (traces not shown above for simplicity but were recorded) was 2.2s (*SD*=0.88s) in the audio condition and 2.7s (*SD*=2.4s) in the haptic condition.

In terms of overall preference, three participants preferred *audio-only* feedback; see Table 2. Reasons included that they were more familiar with audio than haptic signals (P1, P3), and that it was easier to attend to text-to-speech plus audio than to text-to-speech plus haptic (P4). P2's most preferred condition was the *combined* feedback, because she liked to have audio cues for line tracing and haptic cues for start/end of line notifications. In contrast, *haptic-only* feedback was least preferred by three participants. For example, concerned by the desensitization of her nerves, P1 expressed that: *"…if your hands are cold, a real cold air-conditioned room, it's [my tactile sensation] not going to pick it up as well."* P4 also commented on being attuned to sound even in the haptic condition: *"You don't know if it's the top or the bottom [vibrating]…It was the same noise, the same sound."* As shown in Figure 7, ease of use ratings on specific components of the task mirrored overall preference rankings.

	Rank 1	Rank 2	Rank 3
P1	Audio	Combined	Haptic
P2	Combined	Audio	Haptic
P3	Audio	Haptic	Combined
P4	Audio	Combined	Haptic

Table 2: Overall preference rankings per participant. Audio feedback was the most positively received.

	Braille	Screen Reader	Printed Text
P1	3	3	3
P2	3	5	5
P3	4	4	4
P4	5	5	5

Table 3: Ratings comparing prior text reading experiences with HandSight; 1-much worse to 5-much better.

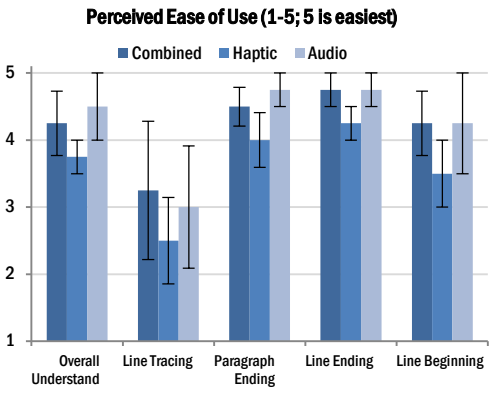


Figure 7: Avg perceived ease of use of different text guidance attributes based on a 5-point scale (1-*very difficult*; 5-*very easy*). Error bars are stderr (*N*=4).

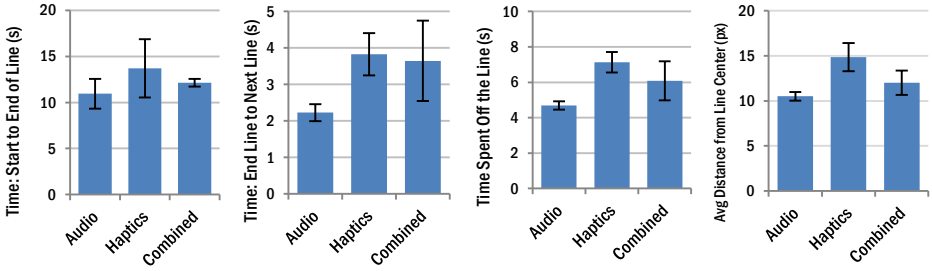


Figure 8: Average performance data from the four user study participants across the three feedback conditions. While preliminary, these results suggest that audio-only feedback may be more effective than the other options tested. Error bars show standard error; (*N*=4).

Participants were also asked to compare their experience with HandSight to braille, screen readers and printed-text reading using 5-point scales (*1-much worse* to *5-much better*). As shown in Table 3, HandSight was perceived to be at least as good (3) or better compared to each of the other reading activities. In general, all participants appreciated HandSight because it allowed them to become more independent when reading non-braille printed documents. For example, P3 stated, *"It puts the blind reading on equal footing with rest of the society, because I am reading from the same reading material that others read, not just braille, which is limited to blind people only"*. P1, who had experience with Optacon [26], Sara CE, and other printed-text scanning devices also commented on HandSight's relative portability.

In terms of performance, we examined four primary measures averaged across all lines per participant (Figure 8): average absolute vertical distance from the line center, time spent off the line (*i.e.*, during which vertical feedback was on), time from start to end of a line, and time from the end of a line to the start of the next line. While it is difficult to generalize based on performance data from only four participants, *audio-only* may offer a performance advantage over the other two conditions. *Audio-only* resulted in the lowest average vertical distance to the line center for all participants. Compared to the *haptic-only* condition, *audio-only* reduced the amount of time spent off the line by about half. It was also faster for all participants than *haptic-only* in moving from the end of a line to the start of the next line. A larger study is needed to confirm these findings and to better assess what impact the feedback conditions have on reading speed from start to end of a line.

5 Discussion

Though preliminary, our research contributes to the growing literature on wearables to improve access to the physical world for the blind (*e.g.*, [12], [19], [21]). The design and initial algorithmic evaluation of our current HandSight prototype show the feasibility of our approach, and highlight important technical issues that we must consider. Additionally, our user study, which evaluated three finger-guidance ap-

proaches using a controlled setup (the iPad test apparatus), found that, in contrast to prior work [21], haptic feedback was the *least* preferred guidance condition. The pitch-controlled audio feedback condition was not only subjectively rated the most preferred but also appeared to improve user performance. Clearly, however, more work is needed. Below, we discuss our preliminary findings and opportunities for future work.

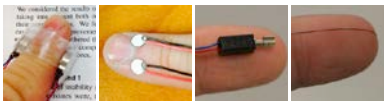


Figure 9: We are evaluating a range of micro-haptic actuators: (a) 10×2.7mm^2 vibro-discs; (b) 5×0.4 mm^2 piezo discs; (c) 3×8 mm^2 vibro-motors; (d) 0.08mm Flexinol wire (shape memory alloy).

Haptic Feedback. Though we have created many different types of finger-mounted haptic feedback in our lab, we tested only one in the user study: when the user moved above or below the current line, s/he would feel a continuous vibration proportional in strength to the distance from the vertical line center. We plan to experiment with form factors, haptic patterns (*e.g.,* intensity, frequency, rhythm, pressure), number of haptic devices on the finger, as well as the type of actuator itself (Figure 9). While our current haptic implementation performed the worst of the feedback conditions, we expect that, ultimately, some form of haptics will be necessary for notifications and finger guidance.

Blind reading. Compared to current state-of-the-art reading approaches, our long-term goals are to: (1) provide more intuitive and precise control over scanning and text-to-speech; (2) increase spatial understanding of the text layout; and (3) mitigate camera framing, focus, and lighting issues. Moreover, because pointing and reading are tightly coupled, finger-based interaction intrinsically supports advanced features such as rereading (for sighted readers, rereading occurs 10-15% of the time [63] and increases comprehension and retainment [64], [65]). We focused purely on reading simple document text, but we plan to investigate more complex layouts so the user can sweep their finger over a document and sense where pictures are located, headings, and so on. We will explore a variety of documents (*e.g.*, plain text, magazines, bills) and household objects (*e.g.*, cans of food, cleaning supplies), and examine questions such as: How should feedback be provided to indicate where text/images are located? How should advanced features such as re-reading, excerpting, and annotating be supported, perhaps, through additional gestural input and voice notes?

Computer Vision. Our preliminary algorithms are efficient and reasonably accurate, but there is much room for improvement. By incorporating constraints on lower-level text features we may be able to rectify vertical perspective effects and affine skew. We can also apply deblurring and image stabilization algorithms to improve the maximum reading speed the system is able to support. Robust and efficient document mosaicking and incorporation of prior knowledge will likely be a key component for supporting a wider range of reading tasks.

Multi-sensory approach. Currently, our prototype relies on only local information gleaned from the on-finger camera. However, in the future, we would like to combine camera streams from both a body-mounted camera (*e.g.,* Orcam [12]) and a finger-mounted camera. We expect the former could provide more global, holistic information about a scene or text which could be used to guide the finger towards a target of interest or to explore the physical document's layout. We could also use the information to improve the performance of the OCR algorithms, by dynamically training the classifier on the page fonts and creating a generative model (*e.g.,* [66]).

6 Conclusion

Our overarching vision is to transform how people with VI access visual information through touch. Though we focused specifically on reading, this workshop paper offers a first step toward providing a general platform for touch-vision applications.

7 References

[1] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010.," *Br. J. Ophthalmol.*, vol. 96, no. 5, pp. 614–8, May 2011.

[2] National Eye Institute at the National Institute of Health, "Blindness Statistics and Data." [Online]. Available: http://www.nei.nih.gov/eyedata/blind.asp. [Accessed: 10-Mar-2014].

[3] D. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 40, no. 1, pp. 25–35, Jan. 2010.

[4] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Wearable real-time stereo vision for the visually impaired," *Eng. Lett.*, vol. 14, no. 2, pp. 6–14, 2007.

[5] J. A. Hesch and S. I. Roumeliotis, "Design and Analysis of a Portable Indoor Localization Aid for the Visually Impaired," *Int. J. Rob. Res.*, vol. 29, no. 11, pp. 1400–1415, Jun. 2010.

[6] A. Hub, J. Diepstraten, and T. Ertl, "Design and Development of an Indoor Navigation and Object Identification System for the Blind," *SIGACCESS Access. Comput.*, no. 77–78, pp. 147–152, Sep. 2003.

[7] R. Manduchi, "Mobile Vision as Assistive Technology for the Blind: An Experimental Study," in *Computers Helping People with Special Needs SE - 2*, vol. 7383, K. Miesenberger, A. Karshmer, P. Penaz, and W. Zagler, Eds. Springer Berlin Heidelberg, 2012, pp. 9–16.

[8] A. Helal, S. E. Moore, and B. Ramachandran, "Drishti: an integrated navigation system for visually impaired and disabled," in *Proceedings Fifth International Symposium on Wearable Computers*, 2001, pp. 149–156.

[9] S. Krishna, G. Little, J. Black, and S. Panchanathan, "A Wearable Face Recognition System for Individuals with Visual Impairments," in *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, 2005, pp. 106–113.

[10] S. Krishna, D. Colbry, J. Black, V. Balasubramanian, and S. Panchanathan, "A Systematic Requirements Analysis and Development of an Assistive Device to Enhance the Social

Interaction of People Who are Blind or Visually," in *Impaired," Workshop on Computer Vision Applications for the Visually Impaired (CVAVI 08), European Conference on Computer Vision ECCV 2008*, 2008.

[11] L. Gade, S. Krishna, and S. Panchanathan, "Person Localization Using a Wearable Camera Towards Enhancing Social Interactions for Individuals with Visual Impairment," in *Proceedings of the 1st ACM SIGMM International Workshop on Media Studies and Implementations That Help Improving Access to Disabled Users*, 2009, pp. 53–62.

[12] OrCam Technologies Ltd, "OrCam - See for Yourself." [Online]. Available: http://www.orcam.com/. [Accessed: 23-Jun-2014].

[13] F. Iannacci, E. Turnquist, D. Avrahami, and S. N. Patel, "The Haptic Laser: Multi-sensation Tactile Feedback for At-a-distance Physical Space Perception and Interaction," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 2047–2050.

[14] V. Khambadkar and E. Folmer, "GIST: A Gestural Interface for Remote Nonvisual Spatial Perception," in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 2013, pp. 301–310.

[15] A. Israr, O. Bau, S.-C. Kim, and I. Poupyrev, "Tactile Feedback on Flat Surfaces for the Visually Impaired," in *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, 2012, pp. 1571–1576.

[16] J. F. Norman and A. N. Bartholomew, "Blindness enhances tactile acuity and haptic 3-D shape discrimination.," *Atten. Percept. Psychophys.*, vol. 73, no. 7, pp. 2323–31, Oct. 2011.

[17] D. Goldreich and I. M. Kanics, "Tactile Acuity is Enhanced in Blindness," *J. Neurosci.*, vol. 23, no. 8, pp. 3439–3445, Apr. 2003.

[18] S. Karim, A. Andjomshoaa, and A. M. Tjoa, "Exploiting SenseCam for Helping the Blind in Business Negotiations," in *Computers Helping People with Special Needs SE - 166*, vol. 4061, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Springer Berlin Heidelberg, 2006, pp. 1147–1154.

[19] S. Nanayakkara, R. Shilkrot, K. P. Yeo, and P. Maes, "EyeRing: A Finger-worn Input Device for Seamless Interactions with Our Surroundings," in *Proceedings of the 4th Augmented Human International Conference*, 2013, pp. 13–20.

[20] 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*, 2012, pp. 147–156.

[21] R. Shilkrot, J. Huber, C. Liu, P. Maes, and N. S. Chandima, "FingerReader: A Wearable Device to Support Text Reading on the Go," *CHI '14 Ext. Abstr. Hum. Factors Comput. Syst.*, no. Vi, pp. 2359–2364, 2014.

[22] F. S. Cooper, J. H. Gaitenby, and P. W. Nye, *Evolution of reading machines for the blind: Haskins Laboratories' research as a case history*. Haskins Laboratories, 1983.

[23] M. Capp and P. Picton, "The optophone: an electronic blind aid," *Eng. Sci. Educ. J.*, vol. 9, no. 3, pp. 137–143, 2000.

[24] E. F. D'Albe, "On a Type-Reading Optophone," *Proc. R. Soc. London. Ser. A*, vol. 90, no. 619, pp. 373–375, 1914.

[25] J. C. Bliss, "A Relatively High-Resolution Reading Aid for the Blind," *Man-Machine Syst. IEEE Trans.*, vol. 10, no. 1, pp. 1–9, Mar. 1969.

[26] D. Kendrick, “From Optacon to Oblivion: The Telesensory Story,” *American Foundation for the Blind AccessWorld Magazine*, vol. 6, no. 4, 2005.

[27] Intel, “Intel Reader.” [Online]. Available: http://www.intel.com/pressroom/kits/healthcare/reader/. [Accessed: 10-Jan-2014].

[28] knfb Reading Technology Inc., “knfb Reader Classic.” [Online]. Available: http://www.knfbreader.com/products-classic.php. [Accessed: 22-Jun-2014].

[29] V. Gaudissart, S. Ferreira, C. Thillou, and B. Gosselin, “SYPOLE: mobile reading assistant for blind people,” in *9th Conference Speech and Computer (SPECOM)*, 2004.

[30] Blindsight, “Text Detective.” [Online]. Available: http://blindsight.com/textdetective/. [Accessed: 01-Nov-2013].

[31] knfb Reading Technology Inc., “kReader Mobile.” [Online]. Available: http://www.knfbreader.com/products-kreader-mobile.php. [Accessed: 24-Jun-2014].

[32] S. Mori, C. Y. Suen, and K. Yamamoto, “Historical review of OCR research and development,” *Proc. IEEE*, vol. 80, no. 7, pp. 1029–1058, Jul. 1992.

[33] H. Shen and J. Coughlan, “Towards a Real-Time System for Finding and Reading Signs for Visually Impaired Users,” in *Computers Helping People with Special Needs SE - 7*, vol. 7383, K. Miesenberger, A. Karshmer, P. Penaz, and W. Zagler, Eds. Springer Berlin Heidelberg, 2012, pp. 41–47.

[34] X. Chen and A. L. Yuille, “Detecting and reading text in natural scenes,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, vol. 2, pp. II–366–II–373 Vol.2.

[35] K. Wang, B. Babenko, and S. Belongie, “End-to-end scene text recognition,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 1457–1464.

[36] C. Jayant, H. Ji, S. White, and J. P. Bigham, “Supporting Blind Photography,” in *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, 2011, pp. 203–210.

[37] R. Manduchi and J. M. Coughlan, “The last meter: blind visual guidance to a target,” in *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI’14)*, 2014, p. To appear.

[38] S. K. Kane, B. Frey, and J. 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*, 2013, pp. 347–350.

[39] OrCam as uploaded by YouTube user Amnon Shashua, “OrCam at Digital-Life-Design (DLD) in Munich,” *Digital-Life-Design*, 2014. [Online]. Available: http://youtu.be/3m9ivtJI6iA?t=2m10s. [Accessed: 22-Jun-2014].

[40] OrCam as uploaded by YouTube user Amnon Shashua, “OrCam TED@NYC,” *TED@NYC*, 2013. [Online]. Available: http://youtu.be/_3XVsCsscyw?t=4m53s. [Accessed: 22-Jun-2014].

[41] A. Crossan and S. Brewster, “Multimodal Trajectory Playback for Teaching Shape Information and Trajectories to Visually Impaired Computer Users,” *ACM Trans. Access. Comput.*, vol. 1, no. 2, pp. 12:1–12:34, Oct. 2008.

[42] B. Plimmer, P. Reid, R. Blagojevic, A. Crossan, and S. Brewster, “Signing on the Tactile Line: A Multimodal System for Teaching Handwriting to Blind Children,” *ACM Trans. Comput. Interact.*, vol. 18, no. 3, pp. 17:1–17:29, Aug. 2011.

[43] J. Su, A. Rosenzweig, A. Goel, E. de Lara, and K. N. Truong, “Timbremap: Enabling the Visually-impaired to Use Maps on Touch-enabled Devices,” in *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2010, pp. 17–26.

[44] S. Harada, H. Takagi, and C. Asakawa, “On the Audio Representation of Radial Direction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 2779–2788.

[45] K. Yatani and K. N. Truong, “SemFeel: A User Interface with Semantic Tactile Feedback for Mobile Touch-screen Devices,” in *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, 2009, pp. 111–120.

[46] K. Yatani, N. Banovic, and K. Truong, “SpaceSense: Representing Geographical Information to Visually Impaired People Using Spatial Tactile Feedback,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 415–424.

[47] N. Noble and B. Martin, “Shape discovering using tactile guidance,” in *Proceeding of the 6th International Conference EuroHaptics*, 2006.

[48] U. Oh, S. K. Kane, and L. Findlater, “Follow that sound: using sonification and corrective verbal feedback to teach touchscreen gestures,” in *Proceedings of the ACM SIGACCESS International Conference on Computers and Accessibility (ASSETS 2013)*, 2013, p. To appear.

[49] S. J. Lederman and R. L. Klatzky, “Hand movements: A window into haptic object recognition,” *Cogn. Psychol.*, vol. 19, no. 3, pp. 342–368, 1987.

[50] K. Johnson, “Neural basis of haptic perception,” in *Stevens’ Handbook of Experimental Psychology: Volume 1: Sensation and Perception*, 3rd Editio., H. Pashler and S. Yantis, Eds. Wiley Online Library, 2002, pp. 537–580.

[51] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay, “Enabling Always-available Input with Muscle-computer Interfaces,” in *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, 2009, pp. 167–176.

[52] D. Morris, T. S. Saponas, and D. Tan, “Emerging input technologies for always-available mobile interaction,” *Found. Trends Human--Computer Interact.*, vol. 4, no. 4, pp. 245–316, 2010.

[53] T. S. Saponas, “Supporting Everyday Activities through Always-Available Mobile Computing,” University of Washington, 2010.

[54] AWAIBA, “NanEye Medical Image Sensors.” [Online]. Available: http://www.awaiba.com/en/products/medical-image-sensors/. [Accessed: 10-Jan-2014].

[55] Samsung, “Samsung Galaxy Gear.” [Online]. Available: http://www.samsung.com/us/mobile/wearable-tech/SM-V7000ZKAXAR. [Accessed: 10-Jan-2014].

[56] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[57] L. Jagannathan and C. Jawahar, “Perspective Correction Methods for Camera Based Document Analysis,” *Proc. First Int. Work. Camera-based Doc. Anal. Recognit.*, pp. 148–154, 2005.

[58] V. Zaliva, “Horizontal Perspective Correction in Text Images,” 2012. [Online]. Available: http://notbrainsurgery.livejournal.com/40465.html.

[59] I. S. Abuhaiba, “Efficient OCR using Simple Features and Decision Trees with Backtracking,” *Arab. J. Sci. Eng.*, 2006, vol. 31, no. 2, pp. 223–244, 2006.

[60] E. Rosten and T. Drummond, “Fusing points and lines for high performance tracking,” in *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, 2005, vol. 2, pp. 1508–1515 Vol. 2.

[61] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features*. Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.

[62] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[63] R. Keefer, Y. Liu, and N. Bourbakis, “The Development and Evaluation of an Eyes-Free Interaction Model for Mobile Reading Devices,” *Human-Machine Syst. IEEE Trans.*, vol. 43, no. 1, pp. 76–91, 2013.

[64] S. L. Dowhower, “Repeated Reading: Research into Practice,” *Read. Teach.*, vol. 42, no. 7, pp. pp. 502–507, 1989.

[65] B. A. Levy, “Text processing: Memory representations mediate fluent reading,” *Perspect. Hum. Mem. Cogn. aging Essays honour Fergus Craik*, pp. 83–98, 2001.

[66] J. Lucke, “Autonomous cleaning of corrupted scanned documents — A generative modeling approach,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3338–3345.

Localization of Skin Features on the Hand and Wrist from Small Image Patches

Lee Stearns[*], Uran Oh[*], Bridget J. Cheng[†], Leah Findlater[*], David Ross[‡], Rama Chellappa[*], Jon E. Froehlich[*]
[*] University of Maryland, College Park, Maryland, United States
[†] Cornell University, Ithaca, New York, United States
[‡] Atlanta VA R&D Center for Vision & Neurocognitive Rehabilitation, Atlanta, Georgia, United States

Abstract*—Skin-based biometrics rely on the distinctiveness of skin patterns across individuals for identification. In this paper, we investigate whether small image patches of the skin can be localized on a user's body, determining not "who?" but instead "where?" Applying techniques from biometrics and computer vision, we introduce a hierarchical classifier that estimates a location from the image texture and refines the estimate with keypoint matching and geometric verification. To evaluate our approach, we collected 10,198 close-up images of 17 hand and wrist locations across 30 participants. Within-person algorithmic experiments demonstrate that an individual's own skin features can be used to localize their skin surface image patches with an F_1 score of 96.5%. As secondary analyses, we assess the effects of training set size and between-person classification. We close with a discussion of the strengths and limitations of our approach and evaluation methods as well as implications for future applications using a wearable camera to support touch-based, location-specific taps and gestures on the surface of the skin.

Keywords— skin texture classification; biometrics and computer vision applications; on-body input

I. Introduction

Previous work in skin classification has largely been in the context of biometrics—that is, determining the *uniqueness* of a user's skin patterns for identification purposes (*e.g.,* [1–7]). In this paper, rather than identifying *who* an image represents, we seek to identify *where* an image is located on a single user's body. More specifically, we investigate to what extent are surface image patches of the hand and wrist localizable?

Being able to determine the location of a small patch of skin could enable a wearable camera to support a range of *on-body interactions*, an emerging paradigm in human-computer interaction (HCI) where users tap or gesture on their own body to control a computing device (*e.g.*, [8–16]). One advantage is that this type of input is always available, allowing the user to, for example, quickly tap or swipe on their palm to answer a phone call or listen to new emails (Figure 1a). On-body interaction is also useful when visual attention is limited because the skin's tactile perception allows for more accurate input than is possible with a touchscreen [17, 18].

Sensing these on-body taps and gestures, however, is a challenging problem. Researchers have investigated a variety of wearable cameras (*e.g.*, [11, 19]) and other sensors (*e.g.*, bio-acoustics [8], ultrasonic rangefinders [9]). While promising, these approaches are limited by the placement and range of the sensor [14, 19], suffer from occlusion [19] or precision [8] problems, or cover the user's skin [15], reducing tactile sensitivity. Instead, we envision using close-up images from a small *finger-mounted camera* (*e.g.*, [20, 21]) to sense and localize user input (Figure 1b). By instrumenting the gesturing finger with a camera, our approach extends the user's interaction space to anything within reach and can support fairly precise location-based input.

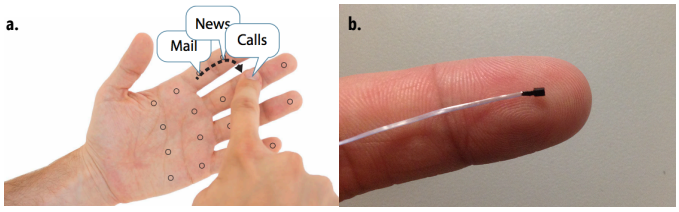


Figure 1. (a) Conceptual visualization of on-hand input to control a mobile phone, as in [17]. (b) Cameras developed for minimally invasive surgeries are small enough to mount on the finger. Shown: AWAIBA NanEye ($1\times1mm^2$, 250×250px resolution) used in [20, 21].

Localizing small (~1–2 cm) image patches within the larger skin surface is similar to partial finger and palm print recognition in forensic applications; however, high-resolution, high-contrast images of ridge impressions are typically needed to reliably extract distinctive point and line features. In contrast, cameras small enough to be mounted on the finger (Figure 1b) are low resolution and low contrast, making it difficult to detect minute ridge features. Several recent biometric systems recognize finger and palm prints using lower-quality images [2, 5–7, 22–27]. Unfortunately, these approaches are frequently designed to align and process the finger or palm image as a whole, and cannot reliably recognize a small portion of the print. To our knowledge, no work has attempted to recognize or localize a small skin patch from live camera images, which we do here.

To ultimately support on-body localization using a finger-mounted camera, we investigate the classifiability of 17 locations on the front and back of the palm, fingers, wrist, nails, and knuckles. We introduce a hierarchical texture classification approach to first estimate the approximate touch location on the body given close-up images of the skin surface and then refine the location estimate using keypoint matching and geometric verification. To evaluate our approach, we collected a skin-surface image dataset consisting of 30 individuals and the 17 hand and wrist locations (10,198 total images).

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617

When testing and training on an individual's own skin data (within-person experiments), our results show that skin patches are classifiable by location under controlled conditions with 96.6% recall and 96.4% precision, suggesting that finger-mounted cameras may be feasible for sensing on-body interactions.

In summary, the contributions of this paper include: (i) a robust algorithmic pipeline for recognizing several different locations on the hand from small patches of skin; (ii) classification results for a dataset consisting of 30 individuals, achieving accuracy above 96% on average for within-person experiments; and (iii) analysis of hand distinctiveness and similarities among users, which may impact accuracy and scalability (*e.g.,* between-person training feasibility).

II. Related Work

Our work applies and extends research in biometrics and on-body interaction, which we describe below.

A. Finger and Palm Biometrics

Work in biometrics has demonstrated that the skin of the hand—specifically, the palm and fingers—contains a large number of highly distinctive visual features that can be used to identify individuals [1, 2]. As noted in the Introduction, we borrow and extend biometrics algorithms, including those for partial fingerprint and palmprint recognition used in forensics [3, 4] as well as techniques that use relatively low resolution and low contrast webcam and mobile phone camera images for person identification and verification [2, 5–7].

Fingerprint and palmprint recognition requires matching an image against a set of stored templates, either directly using the image intensities [22], using texture representations [23, 24], or using minutiae and other point features [3, 4, 6, 27, 28]. A direct match requires precise alignment of the finger or palm as a whole, which is infeasible for the close-up, partial images used in our work. However, we can support partial matching by applying some of the same methods for preprocessing, representing texture, and extracting point features.

To preprocess the images, most approaches apply some filter (*e.g.,* Gabor filters [23]) that enhances the skin's ridges and principal lines. We use a method inspired by Huang *et al.*'s palmprint verification work [29]. To represent texture, Local Binary Pattern (LBP) histograms are a common choice in biometrics [24, 30]. While we explored other texture-based methods, such as Gabor histograms [25] and wavelet principal components [26], we found that they offered negligible improvements over LBP despite their increased computational complexity. To detect point features, some systems use Harris corners [27] or scale-invariant feature transform (SIFT) key points [7, 27]; however, the comparisons between feature descriptors are challenging due to the repetitive nature of finger and palm images. Thus, we instead extract custom features that are based upon Gabor filter response, inspired by [29].

B. On-Body Interaction

A wide variety of wearable sensors have been used to support touch-based input on the user's own body, from arm-worn bio-acoustic sensors [8] and ultrasonic rangefinders [9] to infrared reflectance sensors [14] and touch-sensitive skin overlays [15]; these approaches, however, are limited by occlusion and precision problems, sensor placement and range issues, and/or impeding the user's sense of touch—as noted in the Introduction. Most relevant are approaches that use body-mounted cameras—often depth and/or infrared cameras worn on the chest, shoulder, or head [11–13, 16, 19, 31]. These devices detect on-body gestures from camera images using low-level computer vision techniques such as histogram [19] and motion-based segmentation and tracking [31], and Support Vector Machine (SVM)-based classifiers [31]. For instance, Harrison *et al.*'s [11] *OmniTouch* enables touch-based interactions on a variety of surfaces via a shoulder-mounted depth camera for hand tracking and a pico-projector for visual feedback. While similar to using a finger-mounted camera, body-mounted camera hardware is subject to occlusion issues and limits the interaction space to the *fixed* camera's field of view [19].

In contrast, we envision using a finger-mounted camera that can be freely pointed to interaction targets. Such cameras have been used by *FingerReader* [32] and *HandSight* [21, 33] for reading printed text, and by *Magic Finger* [20], which combines a high-speed optical mouse sensor to track finger movement and a slower but higher-resolution camera to capture detail for texture classification. We were particularly inspired by Magic Finger, which achieved high accuracy (99%) in classifying a set of 22 textures that included table surfaces, white paper, clothing, and two body locations (hand, thumb). The first stage of our surface classification approach is similar to theirs, using an LBP texture representation and an SVM classifier.

III. Touch Localization Pipeline

Robust localization of close-up skin images from a finger-mounted camera is challenging due to the limited field of view (~1–2 cm) and relatively low contrast of the ridges and other skin surface features. To estimate the user's touch location from close-up images, we developed a hierarchical classifier with four stages: (i) preprocessing, (ii) coarse-grained classification, (iii) fine-grained classification, (iv) geometric verification and refinement. The coarse-grained stage classifies an input image into one of five regions: *palm*, *fingers*, *nail*, *knuckle*, and *other* (wrist and back of hand). The fine-grained stage further classifies the image into a discrete location within that region (17 locations in all; see Figures 2 and 5). These locations were selected because previous work has shown that users can reliably locate them with high accuracy even without sight [18]. While our four-stage pipeline integrates multiple known approaches in fingerprint and palmprint enhancement, texture classification, and 2D keypoint matching, our primary innovation is in their novel combination and application towards *localization* rather than identification.

Stage 1: Preprocessing. Images are first preprocessed to remove noise and emphasize ridge features. We apply an efficient median filter [34] to reduce the effect of dirt and other camera noise while preserving the edge information necessary for processing finger and palm prints (Figure 2).

To emphasize the ridgelines, we adapt a technique from Huang *et al.* [29]. However, while they use a modified version of the finite radon transform to emphasize the principal lines and creases of the palm, these features are not as prominent in our images due to the narrow field of view. We instead use Gabor

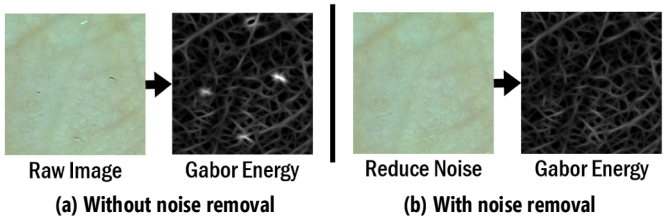


Figure 2. Stage 1 preprocessing first removes dirt and other noise before emphasizing ridge features using the energy of a set of Gabor filters with different orientations. Shown: an example image from the left side of the palm, scaled and cropped to demonstrate the effect that surface artifacts can have on the Gabor energy image.

filters. We compute the Gabor energy image defined as the maximum response at each pixel from a set of Gabor filters with different orientations. Specifically, the energy at pixel location (x, y) is:

$$E_{x,y} = \left|\max_{\theta}\left[G_\theta * (\bar{I}_{x,y} - I_{x,y})\right]\right| \quad (1)$$

where $I_{x,y}$ is the gray-scale pixel value at (x, y) and $\bar{I}_{x,y}$ is the local mean in a window around that location (estimated using a Gaussian smoothing function), G_θ is a discrete Gabor filter with orientation θ, and * is the convolution operator. In our experiments, we use 18 uniformly distributed orientations, with a fixed scale and bandwidth that were chosen empirically based upon the average ridge frequency in our preliminary experiments with a separate set of pilot data. Example energy images are shown in Figure 2, 3, and 4.

Stage 2: Coarse-Grained Classification. After preprocessing, we obtain a rough classification of the image's location using the visual texture, which we represent using LBP histograms. We chose LBP because of its computational efficiency and natural invariance to illumination variations. To improve accuracy and achieve rotation invariance, we use only the uniform patterns alongside the variance of the neighboring values as suggested in [35]. Our implementation uses a 2D histogram with 14 uniform pattern bins and 12 variance bins ($LBP^{riu2}_{12,2}$ and $VAR_{12,2}$, as defined in [35]), computed at 3 scales. These parameters were selected because they provided a balance between classification accuracy and computational efficiency on our pilot data. The histograms for each scale are flattened and concatenated together to produce a 672-element feature vector, which is then normalized. To classify the LBP histograms into coarse-grained body regions, we train a support vector machine (SVM)—commonly used in texture classification (*e.g.*, [20, 22, 36]).

Stage 3: Fine-Grained Classification. We compare the LBP histogram using a template matching approach against *only* the training templates from the coarse-grained region identified in Stage 2. This hierarchical approach reduces the number of possible match locations and enables us to prioritize different features for each region individually (*e.g.*, for the palm we can automatically weight the palmprint texture features that best discriminate the five different palm locations). For template comparisons, we use the χ^2 distance metric, which is known to perform well with LBP histograms (*e.g.*, [37]). Stage 3 produces a sorted list of templates, with the lowest distance representing the most likely match.

Stage 4: Geometric Verification and Refinement. Stage 4 ensures the validity of the texture match and refines the precise touch location using a set of keypoint matching and geometric

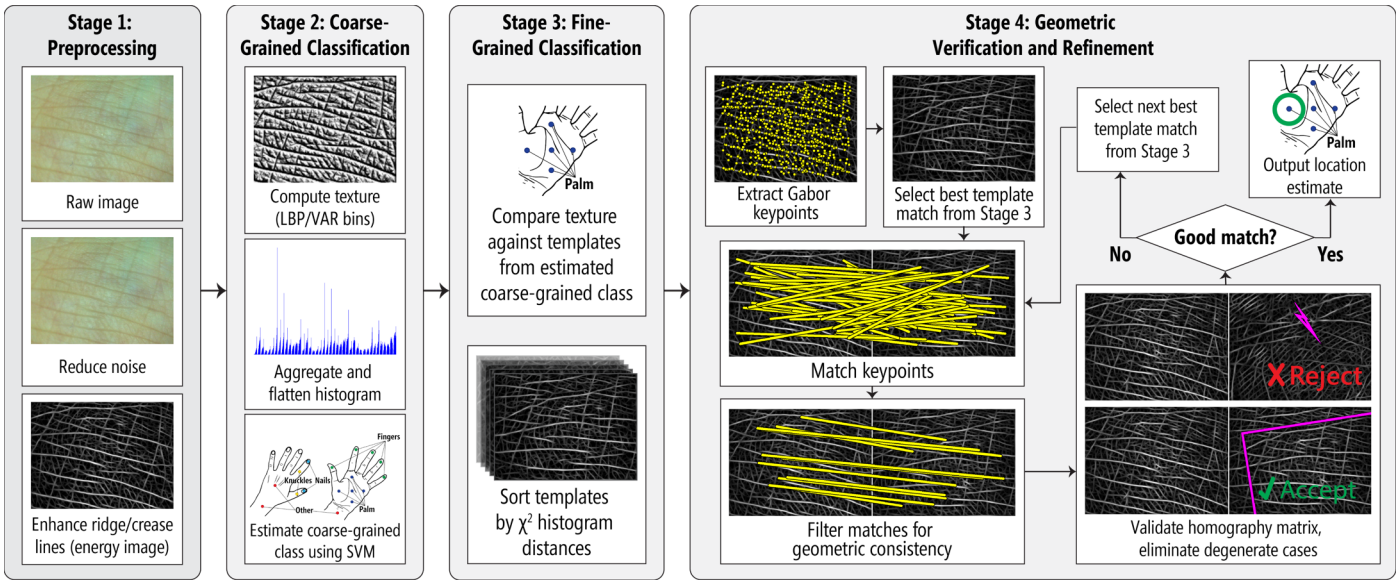


Figure 3. The four stages of our localization algorithm, as applied to an example image from the left side of the palm. First, the image is preprocessed to remove surface artifacts and camera noise before calculating the Gabor energy to emphasize ridge and crease lines. Second, the image is classified into one of five coarse-grained locations (in this case, the palm) using a 2D texture histogram of LBP and pixel variances. Third, the image's texture is compared against the templates from the predicted coarse-grained class, which are sorted by their χ^2 histogram distances to prioritize matching for the next stage. Finally, the image is compared geometrically against images from the predicted coarse-grained class, using a set of custom Gabor keypoints and descriptors. The image is compared against individual templates starting with the most likely match (as predicted in Stage 3), proceeding in order until a template with sufficient geometrically consistent keypoint matches is found. If a geometrically consistent match is found, then the fine-grained location can be estimated with a high degree of certainty (in this case, the left side of the palm); otherwise, the algorithm falls back upon the closest texture match from Stage 3.

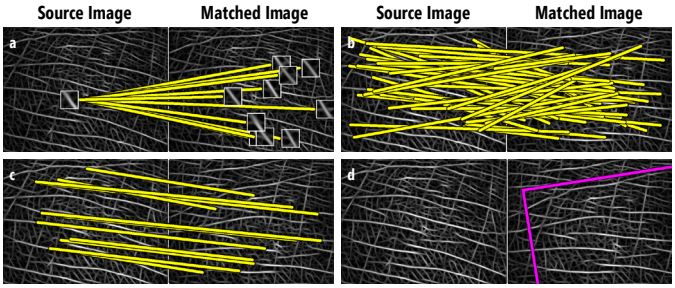


Figure 4. Keypoints in the Gabor energy images frequently appear visually similar (a), leading to a high percentage of mismatches (b). We filter outliers using a series of verification steps to ensure geometric consistency (c and d).

verification steps. We investigated SIFT keypoints [6, 7, 27], Harris corners [27], and fingerprint minutiae [3, 4, 28], but found them too unreliable in preliminary tests. Instead, we use keypoints with a high Gabor filter response at two or more orientations, which tend to lie at the intersections of ridgelines or creases. The Gabor energy values in the 16×16px neighborhood surrounding the keypoint serve as a reliable descriptor. To achieve rotation invariance, we generate multiple descriptors at each keypoint location, rotating the neighborhood for each using the orientation of the filters with locally maximum response strength. We keep a list of keypoints for each training image.

These image patches, however, are frequently visually similar (*e.g.*, Figure 4a), leading to a high percentage of mismatches between the keypoints (Figure 4b). We address this issue using a series of geometric verification steps. First, we filter the matches for orientation consistency, eliminating matches that do not agree with the majority vote for the relative rotation between images (*i.e.*, any more than 20° from the average rotation across all matches). Second, we compute a homography matrix using random sample consensus (RANSAC), identifying inliers and ensuring that there are sufficient geometrically consistent feature matches (*i.e.,* more than the minimum necessary to define a homography; in our experiments, we required 16 consistent matches). Although the palm and fingers are not rigid planar surfaces, in the close-up images we gathered they appear nearly so; we compensate for any irregularities by allowing a greater than usual inlier distance of 10 pixels. Third, we verify that the homography matrix is well behaved using the following constraints, which ensure that the match preserves orientation and does not have extreme variations in scale or perspective:

$$\begin{aligned} &1. \quad H_{11}H_{22} - H_{21}H_{12} > \tfrac{1}{2} \\ &2. \quad \tfrac{1}{2} < \sqrt{{H_{11}}^2 + {H_{21}}^2} < 2 \\ &3. \quad \tfrac{1}{2} < \sqrt{{H_{12}}^2 + {H_{22}}^2} < 2 \\ &4. \quad \sqrt{{H_{31}}^2 + {H_{32}}^2} < \tfrac{1}{1000} \end{aligned} \tag{2}$$

These constraints were selected empirically to eliminate most degenerate cases that could lead to false-positive matches. Fourth and finally, to avoid further degenerate cases, we ensure that the inlier features are not collinear and that they have sufficient spread. We do this by calculating the standard deviation along the two principal directions computed using principal component analysis; if $\sigma_1 < 25$ or $\sigma_1/\sigma_2 > 4$, we declare the match invalid (these numbers were also selected empirically and validated on a separate set of pilot data). If a template match is declared invalid, we proceed to the next best texture match, stopping once we find one that passes all conditions. If a valid match is not found, then we fall back upon the best Stage 3 texture match.

Participant Demographics

Gender	23 female, 7 male	
Age	Mean = 30.6, SD = 11.5, Min = 18, Max = 59	
Race	Black, Afro-Caribbean, or Afro-American	6
	East Asian or Asian-American	5
	Latino or Hispanic American	1
	Non-Hispanic White or Euro-American	14
	South Asian or Indian American	2
	Other or Multiple	2
Palm Size	Mean = 98.3 mm, SD = 10.3 mm, Min = 79.7 mm, Max = 129.5 mm	

Table 1. Our dataset captures variations in gender, age, race, and palm size. Palm size was measured diagonally from the base of the thumb to base of the smallest finger while the fingers were spread and fully extended.

The output of our hierarchical algorithm is an estimated classification of a query image into one of 17 locations, along with a confidence score based upon the texture similarity and the number of inliers for the best template match. From the computed homography matrix, we also obtain a more precise location estimate relative to the matched training templates, potentially enabling finer localization for future explorations.

IV. Data Collection and Dataset

To evaluate our approach, we created an image dataset collected from 30 volunteers (23 female) recruited via campus email lists. The participants were on average 30.6 years old

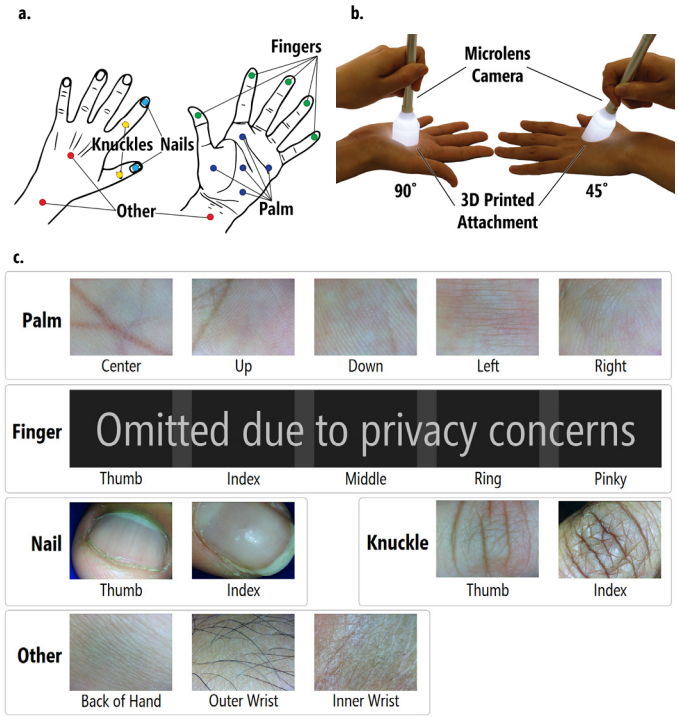


Figure 5. Data collection setup: (a) the 17 close-up image locations on the left hand in 5 coarse-grained regions–coded with different colors; (b) the pen-based camera and physical constraints (one angled at 45° and one at 90°) used for close-up image capture. (c) representative images from our dataset for each of the 17 locations, selected across 12 participants.

(*SD*=11.5, *range*=18–59), and represented a variety of skin tones and palm sizes (Table 1). For each participant, we collected close-up images of 17 locations (Figure 5) using a small 0.3-megapixel micro-lens camera in the shape of a pen.

The micro-lens camera is self-illuminated with a manually adjustable focal length, enabling us to capture clean 640×640px images of the hand from as close as 1cm. We controlled for distance and perspective using two 3D-printed camera attachments that place the camera approximately 2.5cm from the surface of the hand, at either a 90° or 45° angle (Figure 5b). Compared to a finger-mounted camera, this form factor enabled us to more easily control for variables such as distance, perspective, focus, and lighting, while still capturing images that are representative of our target domain. Ultimately we expect to use a smaller camera similar to Figure 1b.

Participants used the camera to point to 17 locations on the hand and palm, with 10 trials for each location and two perspectives (45° and 90°) resulting in 340 images per person. Rather than point 10 times in a row to the same location, the order of trials was randomized to provide natural variation in translation, rotation, and pressure (which impacts scale and focus). In total, we have 10,198 close-up micro-lens images across the 30 participants (one participant skipped two trials). While we would like to release this dataset publicly, we cannot do so without risking the privacy of our participants.

V. Experiments and Results

We first describe results related to coarse- and fine-grained hand classification performance before presenting secondary analyses related to the effect of training sample size on performance and between-person classification. Our analyses report standard measures including precision, recall, and F_1 scores. These metrics are more informative than accuracy due to the uneven number of training examples per class our hierarchy defines.

A. Within-person Classification

To evaluate the overall location-level classifiability of the hand, we conducted a within-person classification experiment. We used an *n*-fold, leave-one-out cross-validation approach. Our results are the average across all 20 folds for each of the 30 participants. We first present aggregate results before examining performance by location and by participant.

Stage 2: Coarse-grained Classification Confusion Matrix

	Palm	*Finger*	*Nail*	*Knuckle*	*Other*
Palm	**99.0%**	0.5%			0.5%
Finger	0.6%	**99.3%**	0.1%		
Nail	0.2%	0.1%	**99.7%**	0.1%	
Knuckle			0.2%	**99.1%**	0.7%
Other	0.6%		0.1%	0.5%	**98.8%**

Table 2: Classification percentages for classes at the coarse-grained level. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row).

At the coarse-grained level (Stage 2), the average precision is 99.1% (*SD*=0.9%) and average recall is 99.2% (*SD*=0.8%). At the fine-grained level (Stage 3), the average precision is 88.2% (*SD*=4.4%) and recall is 88.0% (*SD*=4.5%). After performing geometric validation and refinement (Stage 4), fine-grained classification increases to 96.6% precision (*SD*=2.2%) and 96.4% recall (*SD*=2.3%). The high precision and recall values demonstrate the feasibility of using close-up images to classify locations on the hand and wrist. Stage 2 precision and recall are very high (above 99%), which is important because errors in estimating the coarse-grained region will propagate to the next stage (a limitation of our hierarchical approach). Across all stages, we observed classification errors that were caused primarily by similarities between the locations' visual textures, poor image quality, and insufficient overlap between the training and testing images, although the high accuracies meant that there was not enough data for statistical analysis of the errors.

To examine the impact of different hand/wrist locations on performance, we created confusion matrices for Stage 2 (coarse-grained) and Stage 4 (fine-grained) classifications. See Tables 2 and 3 respectively. The locations with the lowest F_1 score were those on the back of hand (*M*=92.3%; *SD*=10.1%) and wrist (*M*=91.8%; *SD*=8.4%), which appear visually similar (Figure 6). This was true to a lesser extent across all coarse-grained regions, with the textures of different locations within each region appearing similar. While Stage 4 geometric validation reduced misclassifications, it was not always successful. For example, in some cases, an image for a participant did not sufficiently overlap any other image in the dataset, preventing

Stage 4: Fine-grained Classification Confusion Matrix

	Palm					*Fingers*					*Nails*		*Knuckles*		*Other*		
	C	*U*	*D*	*L*	*R*	*1st*	*2nd*	*3rd*	*4th*	*5th*	*1st*	*2nd*	*1st*	*2nd*	*BH*	*OW*	*IW*
Palm Center (C)	**98.3%**					0.2%		0.2%		0.2%					0.2%	0.5%	0.5%
Palm Up (U)	0.2%	**98.5%**		0.2%	0.2%		0.2%			0.3%						0.2%	0.3%
Palm Down (D)	0.3%	1.2%	**95.7%**	0.2%	1.7%	0.3%	0.2%						0.2%			0.2%	0.2%
Palm Left (L)	0.3%	0.3%	0.2%	**98.7%**	0.3%	0.3%											0.7%
Palm Right (R)	0.7%	0.5%	0.3%	0.5%	**97.5%**	0.2%	0.2%			0.2%							
1st Finger		0.5%	0.2%	0.2%	0.7%	**96.3%**	0.3%	0.5%	0.5%	0.7%					0.2%		
2nd Finger		0.3%			0.2%	0.3%	**95.8%**	1.7%	0.5%	1.2%							
3rd Finger			0.3%			0.2%	1.3%	**95.4%**	2.2%	0.7%							
4th Finger			0.2%				0.3%	1.8%	**95.3%**	2.3%							
5th Finger		0.2%		0.2%			0.3%	0.5%	1.5%	**97.0%**	0.3%						
1st Nail				0.2%							**98.2%**	1.7%					
2nd Nail			02%						0.2%		0.5%	**99.0%**		0.2%			
1st Knuckle											0.2%		**97.3%**	1.2%	0.2%	0.2%	1.0%
2nd Knuckle												0.2%	0.8%	**98.8%**		0.2%	
Back of Hand (BH)				0.2%									0.5%	0.2%	**92.2%**	4.7%	2.3%
Outer Wrist (OW)	0.2%												0.2%		6.0%	**90.2%**	3.5%
Inner Wrist (IW)	0.7%	0.7%	0.2%	0.2%	0.3%						0.2%		0.7%	0.2%	0.8%	0.5%	**96.2%**

Table 3: Classification percentages for classes at the fine-grained level (Stage 4 output), averaged across 20 trials and 30 participants. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row).

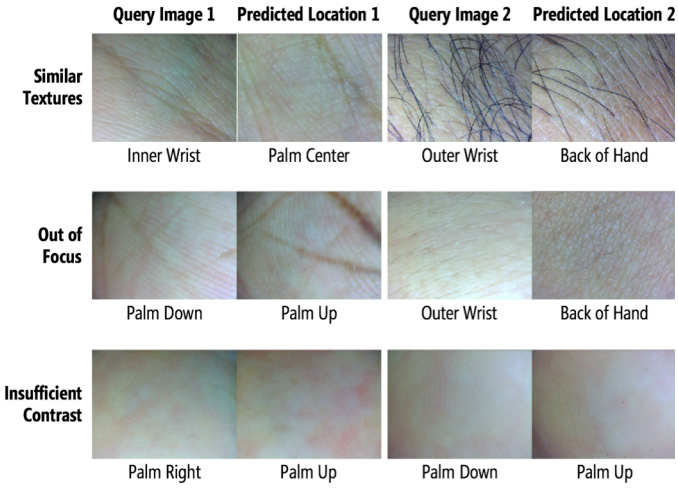


Figure 6: Classification errors were caused primarily by similarities between the locations' visual textures and poor image quality. Each set of images shows, in order, two examples (from different participants) of an incorrectly classified test image along with a training image from the predicted location.

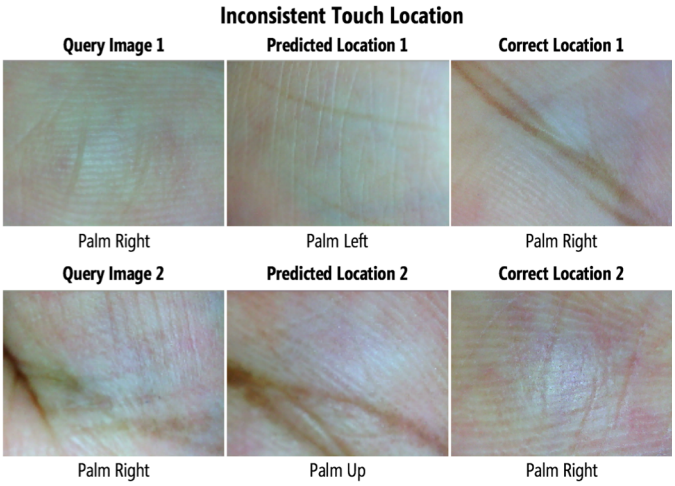


Figure 7: Classification errors for several participants were also caused by inconsistent touch locations. Shown are two examples (from two different participants) where the locations were far enough apart to appear as entirely unrelated images.

geometric keypoint matching (Figure 7). In these cases the algorithm fell back to the best Stage 3 texture match.

To examine how performance varies across individuals, Figure 8a shows F_1 scores broken down by participant. F_1 scores ranged from 95.9% to 100.0% at the coarse-grained level (Stage 2) and 86.5% to 99.7% at the fine-grained level (Stage 4). Participant 29 performed the worst, with a Stage 4 F_1 score of 86.5%—4.4 standard deviations below the mean. Based on a qualitative examination, we found decreased skin contrast with fewer distinctive finger and palm features, as well as significant variations in translation, rotation, and image focus for each location. In comparison, the top performing participants had high contrast skin textures, more consistent pressure (resulting in fewer variations in lighting and focus), and greater consistency in returning to the same touch location each trial.

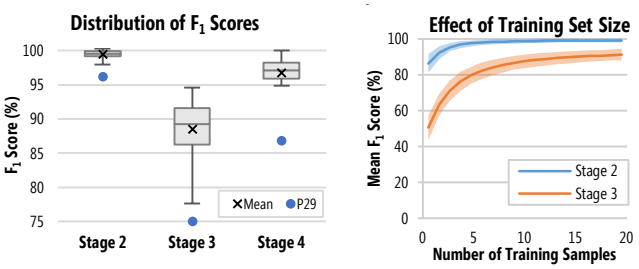


Figure 8: (a) Distribution of F_1 scores by participant, with outlier P29 marked by the blue dot; (b) Effect of the number of training examples on mean texture classification F1 score at coarse-grained (Stage 2, blue) and fine-grained (Stage 3, orange) levels.

Our dataset captured some variations in age, skin tone, and hand size, although not in large enough numbers to confidently determine the effect those factors may have on classification performance (see Table 1). We found no significant correlations across our 30 participants with any of these variables; however further work with a larger, more diverse participant pool is needed.

B. Effect of Training Set Size on Performance

To explore performance as a function of training set size, we tested our algorithms again using *n*-fold cross-validation but this time varying the number of training samples from $m = 1$ to 19. Specifically, we randomly selected from the 20 images per class available for each participant, with one image set aside for testing. Figure 8b shows the average texture classification accuracy at the coarse-grained (Stage 2) and fine-grained (Stage 3) levels when increasing the number of training examples. To reduce the effect of selecting the images randomly and obtain a more representative estimate, we averaged the results of 10 randomized trials. Each point represents the average F_1 score across all participants, locations, and trials when trained using *m* examples. Accuracy begins to level off above five training images per location, especially at the coarse-grained level (which approaches 100% accuracy). However, performance at both levels steadily improves as the number of training images is increased. We did not evaluate Stage 4 for this experiment as its performance depends largely upon the amount of spatial overlap between the training and testing images rather than the number of training samples.

C. Between-person Classification

To potentially bootstrap the training set and to identify similarities across individuals, we conducted a secondary classification experiment in which the training set and testing set consisted of images from different participants (*i.e.,* between-person experiments). More specifically, we employed *n*-fold cross-validation, where each fold trained on data from 29 participants and tested on the remaining participant. We did not expect this approach to yield a high accuracy, especially at the fine-grained level since finger and palm prints can vary significantly person to person (which is the basis of biometric identification). However, we hoped to discover textural similarities across participants that could be used to boost future classifiers to either improve accuracy or reduce the amount of per user training.

Between-person Coarse-grained Classification Confusion Matrix

	Palm	*Finger*	*Nail*	*Knuckle*	*Other*
Palm	**55.2%**	16.8%	7.8%	4.0%	38.8%
Finger	8.1%	**85.5%**	10.4%	2.1%	2.3%
Nail	0.2%	3.4%	**85.3%**	4.4%	0.9%
Knuckle	1.2%	0.2%	1.3%	**67.8%**	18.2%
Other	12.4%	4.1%	0.1%	18.2%	**60.3%**

Table 4: Between-person classification percentages for classes at the coarse-grained level. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row).

As expected, the between-person classification results are lower than the within-person results. At the coarse-grained level, our classification algorithms achieve an average precision of 72.6% (*SD*=12.9%) and recall of 70.8% (*SD*=12.3%). Still, these results are considerably higher than chance for five classes (20%) or majority-vote for the palm class (29.4%). See Table 4 for a confusion matrix. Average precision at the fine-grained level is 27.1% (*SD*=7.5%) and recall of 26.1% (*SD*=5.8%), which are also well above chance for 17 classes (5.9%). Although these accuracies are clearly too low to support a reliable user interface without an individual training procedure, they may provide enough classification information to allow for bootstrapping.

VI. Discussion

Our controlled experiments explored the distinguishability of small image patches on the surface of the hand and wrist for localization purposes. In our within-person experiments we were able to achieve an average F_1 score above 99% at the coarse-grained level (Stage 2) and above 96% at the fine-grained level (Stage 4), which suggests that skin-surface image patches can be classified and localized on the body with high levels of accuracy. While an end-to-end deep learning approach may be more elegant, our more heuristic approach requires substantially less training data, and our performance results suggest that an on-body input system applying our algorithms is feasible. Here, we reflect on the implications of our findings as well as challenges for implementing a real-time system.

A. Expanding On-Body Input

While we only evaluated locations on the hand and wrist, our finger-mounted approach should support a range of input locations within the user's reach, including on-body and off-body surfaces (*e.g.*, tabletops). This is in contrast to most previous on-body input approaches that are more limited by their fixed sensor placements and range. Although recognition accuracy may drop as the number of locations increases (*e.g.*, thigh, forearm), we expect to boost performance through improvements to our hierarchical approach. Performance was particularly high at the first level of the hierarchy, with an F_1 score above 99%. Thus, for each region we could apply different preprocessing and matching approaches at the second level that are tuned specifically to distinguish the fine-grained locations within that region. For example, we could extract knuckle-specific features (*e.g.*, [2]) to distinguish knuckle locations, which may require completely different algorithms than the palm locations. Similarly, it will be important to explore the feasibility of extending the localization hierarchy further, for regions that can support an even finer level of granularity beyond the locations studied (*e.g.*, palm, fingers); such granularity could enable highly precise on-body interactions (*e.g.*, sliding your finger along your palm to trace a map route).

B. Training a Camera-Based On-Body Localization System

The procedure for training a new user may impact both algorithmic performance and user perceptions toward the system. As shown in Figure 7b, classification performance improves with the number of training examples, but begins to level off after five examples per class. However, it may be possible to boost accuracy while simultaneously reducing the number of training examples that are required of a new user. The images in our dataset relied on natural variations that were introduced through randomization during data collection. To potentially improve performance, the training interface could prompt the user to vary rotations, poses, and perspectives–similar to Apple's iPhone training procedure for their fingerprint sensor. In addition, as our preliminary experiments indicate, it may be possible to bootstrap the system using between-person data and reduce the amount of training required for a new user. This approach would work especially well in our first stage of classification to recognize surface classes that appear similar across many users (*e.g.*, skin, knuckles, clothing).

C. Limitations and Future Work

Our experiments were conducted under controlled conditions, but a real-time system would likely need to deal with greater variations in image quality. Although we randomized trial order to introduce natural variation in translation, rotation, and pressure, we carefully controlled for other variables such as distance, lighting, and perspective. A finger-worn camera will likely constrain this complexity, potentially mitigating these concerns. For example, distance will remain relatively constant during touch-based interactions since the camera can be positioned at a fixed location on the finger and lighting can be controlled via a self-illuminated camera. While perspective may vary considerably, our results show that our algorithm functions well for both 90-degree and 45-degree perspectives. Further work is necessary to explore variations under less controlled conditions, including potential changes over time (*e.g.*, due to differences in humidity/dryness), as well as other variations in skin surface textures and features due to age, skin tone, and hand size. The above mitigating factors suggest that our approach should still be applicable.

Our work focused solely on RGB camera-based sensing using static images. Future research should explore other imaging and non-imaging sensors as well as combining video and multiple sensor streams (sensor fusion). For example, hyperspectral imaging would expose veins and other sub-dermal features that could be used for localization as well as improve the contrast of surface features across a wider range of skin tones (*e.g.*, [38]). Depth sensors could provide 3D geometry of the hand and ridges, potentially improving robustness to variations in perspective and allowing us to more reliably extract finger and palm print features to use for localization (*e.g.*, [39]). Finally, non-imaging sensors (*e.g.,* infrared reflectance [14] or inertial motion [40]) could provide complementary information to help resolve visual ambiguities and better integrate localization with gesture recognition.

VII. Conclusion

This paper introduces an algorithmic pipeline for recognizing low-resolution, close-up images of several different locations on the hand/wrist with an average F_1 score of 96.5% for within-person skin patch classification. While future work will need to address potential implementation challenges with a real-time system, our results suggest that a finger-mounted computer vision approach to support location-based on-body interaction should be feasible and that the system training process may be able to be bootstrapped using a dataset of hand images collected from multiple individuals.

References

[1] A. Meraoumia, S. Chitroub, and A. Bouridane, "Fusion of Finger-Knuckle-Print and Palmprint for an Efficient Multi-Biometric System of Person Recognition," in *Proc. of ICC '11*, 2011, pp. 1–5.

[2] M. Choraś and R. Kozik, "Contactless palmprint and knuckle biometrics for mobile devices," *Pattern Anal. Appl.*, vol. 15, no. 1, pp. 73–85, Feb. 2012.

[3] A. K. Jain and J. Feng, "Latent Palmprint Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1032–1047, Jun. 2009.

[4] Eryun Liu, A. K. Jain, and Jie Tian, "A Coarse to Fine Minutiae-Based Latent Palmprint Matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2307–2322, Oct. 2013.

[5] M. O. Derawi, B. Yang, and C. Busch, "Fingerprint Recognition with Embedded Cameras on Mobile Phones," in *Security and Privacy in Mobile Info. and Com. Sys.*, no. Jan 2012, Springer, 2012, pp. 136–147.

[6] A. Morales, M. A. Ferrer, and A. Kumar, "Improved palmprint authentication using contactless imaging," in *IEEE Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2010, pp. 1–6.

[7] X. Wu, Q. Zhao, and W. Bu, "A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors," *Pattern Rec.*, vol. 47, no. 10, pp. 3314–3326, Oct. 2014.

[8] C. Harrison, D. Tan, and D. Morris, "Skinput: Appropriating the Body As an Input Surface," in *Proceedings of CHI '10*, 2010, pp. 453–462.

[9] R.-H. Liang, S.-Y. Lin, C.-H. Su, K.-Y. Cheng, B.-Y. Chen, and D.-N. Yang, "SonarWatch: Appropriating the Forearm as a Slider Bar," in *SIGGRAPH Asia 2011 Emerging Technologies*, 2011, p. 5.

[10] N. Dezfuli, M. Khalilbeigi, J. Huber, F. Müller, and M. Mühlhäuser, "PalmRC: Imaginary Palm-Based Remote Control for Eyes-free Television Interaction," in *Proc. of EuroITV '12*, 2012, p. 27.

[11] C. Harrison and A. D. Wilson, "OmniTouch: wearable multitouch interaction everywhere," in *Proc. of UIST '11*, 2011, pp. 441–450.

[12] E. Tamaki, T. Miyaki, and J. Rekimoto, "Brainy Hand: an earworn hand gesture interaction device," in *Extended Abstracts of ACM CHI 2009*, 2009, pp. 4255–4260.

[13] P. Mistry and P. Maes, "SixthSense: A wearable gestural interface," in *Proc. of ACM SIGGRAPH Asia*, 2009, p. Article No. 11.

[14] M. Ogata, Y. Sugiura, Y. Makino, M. Inami, and M. Imai, "SenSkin: Adapting Skin as a Soft Interface," in *Proc. UIST '13*, 2013, pp. 539–544.

[15] M. Weigel, T. Lu, G. Bailly, A. Oulasvirta, C. Majidi, and J. Steimle, "iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing," in *Proc. CHI'15*, 2015, pp. 2991–3000.

[16] U. Oh and L. Findlater, "Design of and subjective response to on-body input for people with visual impairments," in *Proc. of ASSETS '14*, 2014, pp. 115–122.

[17] S. G. Gustafson, B. Rabe, and P. M. Baudisch, "Understanding Palm-based Imaginary Interfaces: The Role of Visual and Tactile Cues when Browsing," in *Proc. of CHI '13*, 2013, pp. 889–898.

[18] U. Oh and L. Findlater, "A Performance Comparison of On-Hand versus On-Phone Non-Visual Input by Blind and Sighted Users," *ACM Trans. Access. Comput.*, vol. 7, no. 4, p. 14, 2015.

[19] S. Gustafson, C. Holz, and P. Baudisch, "Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device," in *Proc. of UIST '11*, 2011, pp. 283–292.

[20] X.-D. Yang, T. Grossman, D. Wigdor, and G. Fitzmaurice, "Magic finger: always-available input through finger instrumentation," in *Proc. of UIST '12*, 2012, pp. 147–156.

[21] L. Stearns, R. Du, U. Oh, Y. Wang, R. Chellappa, L. Findlater, and J. E. Froehlich, "The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind," *Proc. ECCV '14, Workshop on Assistive Computer Vision and Robotics*, 2014.

[22] J. Doublet, M. Revenu, and O. Lepetit, "Robust GrayScale Distribution Estimation for Contactless Palmprint Recognition," in *IEEE Conference on Biometrics: Theory, Applications, and Systems 2007*, 2007, pp. 1–6.

[23] V. Kanhangad, A. Kumar, and D. Zhang, "A Unified Framework for Contactless Hand Verification," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3, pp. 1014–1027, Sep. 2011.

[24] G. K. Ong Michael, T. Connie, and A. B. Jin Teoh, "Touch-less palm print biometrics: Novel design and implementation," *Image Vis. Comput.*, vol. 26, no. 12, pp. 1551–1560, 2008.

[25] Wai Kin Kong and D. Zhang, "Palmprint texture analysis based on low-resolution images for personal authentication," in *Proc. Pattern Recognition '02*, 2002, vol. 3, pp. 807–810.

[26] M. Ekinci and M. Aykut, "Palmprint Recognition by Applying Wavelet-Based Kernel PCA," *Comput. Sci. Technol.*, vol. 23, no. 107, pp. 851–861, 2008.

[27] A. S. Parihar, A. Kumar, O. P. Verma, A. Gupta, P. Mukherjee, and D. Vatsa, "Point based features for contact-less palmprint images," in *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, 2013, pp. 165–170.

[28] M. Laadjel, A. Bouridane, F. Kurugollu, O. Nibouche, and W. Yan, "Partial Palmprint Matching Using Invariant Local Minutiae Descriptors," in *Transactions on Data Hiding and Multimedia Security*, 2010, pp. 1–17.

[29] D.-S. Huang, W. Jia, and D. Zhang, "Palmprint verification based on principal lines," *Patt. Recog.*, vol. 41, no. 4, pp. 1316–1328, Apr. 2008.

[30] L. Nanni and A. Lumini, "Local binary patterns for a hybrid fingerprint matcher," *Patt. Recog.*, vol. 41, no. 11, pp. 3461–3466, Nov. 2008.

[31] Harrison, S. Ramamurthy, and Hudson, "On-body interaction: armed and dangerous," *Proc. TEI '12*, pp. 69–76, 2012.

[32] R. Shilkrot, J. Huber, M. E. Wong, P. Maes, and S. Nanayakkara, "FingerReader: a wearable device to explore printed text on the go," in *Proc. of CHI '15*, 2015, pp. 2363–2372.

[33] L. Findlater, L. Stearns, R. Du, U. Oh, D. Ross, R. Chellappa, and J. E. Froehlich, "Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch," in *Proc. of ASSETS '15*, 2015, pp. 383–384.

[34] K. Kanagalakshmi and E. Chandra, "Performance evaluation of filters in noise removal of fingerprint image," in *2011 3rd International Conf. on Electronics Computer Technology*, 2011, vol. 1, pp. 117–121.

[35] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[36] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition," *Pattern Recognition*, vol. 42, no. 7, pp. 1408–1418, 2009.

[37] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns," in *Proc. of ECCV '04*, 2004, pp. 469–481.

[38] M. Goel, S. N. Patel, E. Whitmire, A. Mariakakis, T. S. Saponas, N. Joshi, D. Morris, B. Guenter, M. Gavriliu, and G. Borriello, "HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications," in *Proc. of UbiComp '15*, 2015, pp. 145–156.

[39] W. Li, D. Zhang, L. Zhang, G. Lu, and J. Yan, "3-D Palmprint Recognition With Joint Line and Orientation Features," *IEEE Trans. Syst. Man, Cybern. Part C*, vol. 41, no. 2, pp. 274–279, Mar. 2011.

[40] L. Jing, Y. Zhou, Z. Cheng, and T. Huang, "Magic ring: A finger-worn device for multiple appliances control using static finger gestures," *Sensors*, vol. 12, no. 5, pp. 5775–5790, 2012.

Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras

LEE STEARNS, University of Maryland, College Park
RUOFEI DU, University of Maryland, College Park
URAN OH, University of Maryland, College Park
CATHERINE JOU, University of Maryland, College Park
LEAH FINDLATER, University of Maryland, College Park
DAVID A. ROSS, Atlanta VA R&D Center of Excellence in Vision and Neurocognitive Rehabilitation (CVNR)
JON E. FROEHLICH, University of Maryland, College Park

The recent miniaturization of cameras has enabled finger-based reading approaches that provide blind and visually impaired readers with access to printed materials. Compared to handheld text scanners such as mobile phone applications, mounting a tiny camera on the user's own finger has the potential to mitigate camera framing issues, enable a blind reader to better understand the spatial layout of a document, and provide better control over reading pace. A finger-based approach, however, also introduces the need to guide the reader in physically navigating a document, such as tracing along lines of text. While previous work has proposed audio and haptic directional finger guidance for this purpose, user studies of finger-based reading have not provided an in-depth performance analysis of the finger-based reading process. To further investigate the effectiveness of finger-based sensing and feedback for reading printed text, we conducted a controlled lab experiment with 19 blind participants, comparing audio and haptic directional finger guidance within an iPad-based testbed. As a small follow-up, we later asked 4 of those participants to return and provide feedback on a preliminary wearable prototype called HandSight. Findings from the controlled experiment show similar performance between haptic and audio directional guidance, although audio may offer an accuracy advantage for tracing lines of text. Subjective feedback also highlights tradeoffs between the two types of guidance, such as the interference of audio guidance with speech output and the potential for desensitization to haptic guidance. While several participants appreciated the direct access to layout information provided by finger-based exploration, important concerns also arose about ease of use and the amount of concentration required. We close with a discussion on the effectiveness of finger-based reading for blind users and potential design improvements to the HandSight prototype.

General Terms: Human Factors

Additional Key Words and Phrases: Accessibility, Visual Impairments, Wearables, Real-Time OCR

1. INTRODUCTION

Despite the increased availability of digital information and screen reader software, reading printed text materials remains an important but challenging task for people who are blind or visually impaired. The inability to read menus, receipts and handouts, bills and other mail can negatively impact the daily activities of those living with visual impairment (*e.g.*, [Brady et al. 2013; Haymes et al. 2002]). Although braille has long provided a promising alternative, fewer than 10% of the approximately 2 million adults with severe visual impairment in the United States are braille literate [National Federation of the Blind Jernigan Institute 2009; National Center for Health Statistics 2012], and many materials are not available in braille format.

Although many devices and mobile applications—such as *SARA CE*[1], *KNFB Reader iOS*[2], and *OrCam*[3]—attempt to provide access to printed materials through camera capture and optical character recognition (OCR), open questions remain. One

[1] SARA CE: http://www.freedomscientific.com/Products/LowVision/SARA

[2] KNFB Reader: http://knfbreader.com/

[3] OrCam: http://www.orcam.com/



(a) Camera Sensor (b) Finger-mounted Camera (c) Full Prototype

Figure 1: HandSight uses a 1×1mm² AWAIBA NanEye 2C camera developed for minimally invasive surgeries (e.g., endoscopies) that can capture 250×250px images at 44fps (a). Also shown are two early designs for our finger-based reading system (b) and (c). Future designs can be made much smaller.

challenge is how to help blind readers properly aim the camera so that a target object is completely visible and centered within the camera's field of view (e.g., [Vázquez and Steinfeld 2012; Jayant et al. 2011; Cutter and Manduchi 2015]). To accommodate this issue, the popular KNFB Reader iPhone application, for example, provides a spoken report to describe whether the document is fully visible and rotated correctly. Another challenge is how to interpret and communicate documents with complex layouts such as newspapers or menus. Determining which blocks of text to read, in what order, and what layout details to convey are known issues even with digital content [Lazar et al. 2007; Borodin et al. 2010].

As an alternative to these handheld solutions, recent advances in camera technology have led to new wearable approaches that augment the reader's finger with computer vision capabilities [Shilkrot et al. 2014; Shilkrot et al. 2015; Stearns et al. 2014; Yang et al. 2012]. Specifically for reading printed materials, *FingerReader* [Shilkrot et al. 2014; Shilkrot et al. 2015] and our past work [Stearns et al. 2014] both explore ring-based devices with an embedded camera that allows blind readers to trace their finger over printed text and hear real-time speech output. Compared to mobile applications, this finger-based approach may mitigate overhead camera framing issues, enable a blind reader to better understand the spatial layout of a document, and provide better control over pace and rereading.

A finger-based approach, however, also introduces new challenges that have not been fully investigated. Because the field of view from a finger-mounted camera is limited, the reader must precisely trace along the current line of text so that the image does not get cut off or distorted. Physical navigation through the document is also needed to support reading, such as finding the start of a text passage and moving from one line to the next. Thus, a finger-based reading approach is contingent not only on accurate text capture and OCR, but also on effective finger guidance. To date, however, user studies of a finger-mounted approach have focused on feasibility with small sample sizes (3-4 participants) and have not reported on quantitative performance metrics [Stearns et al. 2014; Shilkrot et al. 2014; Shilkrot et al. 2015]. This prevents an in-depth understanding of finger guidance effectiveness, reading performance, and user reactions. The most recent of these studies underscores the need for further investigation: despite the theoretical advantages of finger-based reading, all three participants found it difficult to read text with FingerReader [Shilkrot et al. 2015]. This provokes the question: *why*? To what extent are finger-based cameras a viable accessibility solution for reading printed text? What design choices can improve this viability?

To further investigate the feasibility of a finger-based sensing and feedback system for reading printed text, we conducted a controlled lab experiment to compare audio and haptic directional finger guidance with 19 blind participants using an iPad-based testbed (Study I). The primary goal was to compare the effects of the two guidance methods in terms of line tracing accuracy, reading speed, comprehension (through standardized comprehension questions), and subjective response. We later also randomly selected 4 of those participants to provide feedback on a preliminary wearable prototype, called *HandSight*, so as to help guide its design (Study II). HandSight is an extension of our previous work [Stearns et al. 2014]. These participants also provided feedback on the use of a smartphone app (KNFB Reader iOS) to read printed documents, which allowed us to compile a list of some of the relative advantages and disadvantages of each. An overview of preliminary results from these studies was presented in an ASSETS poster [Findlater et al. 2015].

The findings from Study I showed similar performance between haptic and audio directional guidance, although audio may offer an accuracy advantage for line tracing. While a small majority of participants preferred haptic guidance to audio, the overall split reflects contradictions found in previous research [Stearns et al. 2014; Shilkrot et al. 2014; Shilkrot et al. 2015]. Open-ended comments also highlight the tradeoffs of the two types of guidance, such as the interference of audio guidance with speech output and the potential for desensitization to haptic guidance. Finally, while several participants appreciated the direct access to layout information provided with HandSight's exploration mode, and the lower learning curve of HandSight as compared to braille, important concerns arose about ease of use and the amount of concentration required. In the follow-up sessions (Study II), while not offering a controlled comparison, participants appreciated that HandSight provided immediate access to text content without the need to worry about first capturing the document, but overall they preferred the fast and smooth text-to-speech output of KNFB Reader iOS. Combined, these findings lead to new questions about finger-based reading, who may benefit the most from such an approach, and how to refine the design tested in our study.

The contributions of this paper are: (1) empirical results comparing audio and haptic directional finger guidance for a reading task in terms of user performance and subjective response; (2) the implementation and preliminary evaluation of a real-time proof-of-concept system that combines a small finger-mounted camera and feedback mechanism with efficient computer vision algorithms to read printed text; and (3) design reflections for finger-based reading devices for people who are blind. While our long-term goal is to investigate the many interactions made possible by collocating sensing and feedback on the fingers, our focus here is on the interactions necessary to use such a system to explore and read a physical document.

2. RELATED WORK

We survey existing reading approaches, including braille and dedicated reading devices for people who are blind, and research on enabling access to printed text materials.

2.1. Braille

Braille is one of the oldest and most widely used tactile writing systems for people who are blind [Millar 2003]. Raised line scripts (*e.g.* Moon) and alternative coded dot or dash scripts (*e.g.* Fishburne) are used occasionally for teaching newly blind adults

or for labeling personal items, but none have been as widespread or successful as braille. Braille literacy has been shown to correlate with stronger reading skills and higher employment rates, especially if taught at an early age [Ryles 1996]. However, braille literacy is steadily on the decline. In the past 40 years, the percentage of legally blind school-age children living in the United States who can read or are learning braille has decreased from more than 50% to fewer than 10% [National Federation of the Blind Jernigan Institute 2009]. Moreover, braille is difficult to learn and relatively slow to read—the average reading speed is 125 words per minute (wpm), as opposed to 250-300 wpm for sighted readers [Millar 2003] and even faster for experienced screen reader users [Asakawa et al. 2003]. Additionally, many physical materials still exist that are not available in digital or braille formats, and improving the accessibility of these materials could substantially improve the quality of life of blind individuals [Brady et al. 2013; Haymes et al. 2002].

Braille is a spatial medium, in much the same way as a printed page. Sighted readers use spatial information to help locate specific regions within a document (*e.g.,* paragraphs, headings, captions, sidebars) [Keefer et al. 2013], and evidence suggests that braille readers do too [Millar 2003]. In braille documents, text is organized in straight lines with indentations to identify distinct blocks of text (*i.e.* paragraphs). Some forms also include embossed illustrations, maps, or graphs, with braille letters printed alongside them for labels and captions. These important spatial attributes, however, are not easily conveyed by a screen reader [Di Blas et al. 2004; Kouroupetroglou and Tsonos 2008].

Braille has other advantages as well. As a touch-based system, readers can easily control their reading speed and reread words or lines of text. The ability to reread text is critical, as it can increase comprehension and retention [Dowhower 1989; Levy 2001]. Additionally, although reading styles vary by individual, braille readers who use both hands have been shown to read much more quickly than one-handed readers [Burklen 1932; Millar 2003]. The reason for this is unclear—some individuals primarily use one finger to recognize and verify braille patterns and the other only for tracking position within the page, while others appear to use both fingers in an "enlarged perceptual window" [Millar 2003].

These strengths of braille inspired HandSight's design. As with braille, HandSight should allow blind readers to gain an understanding of the location of document elements (margins, text, image placement), as well as to control reading speed and rereading. HandSight also relies on bimanual operation: the left hand tracks a document's margin while the right hand scans and reads the printed text.

2.2. Dedicated Reading Devices

Scientists have long sought to support blind people in reading printed text by developing new technologies (for reviews: [Cooper et al. 1983; Capp and Picton 2000]). Many early so-called "reading machines for the blind" used a sensory substitution approach where the visual signals of words were converted to non-verbal auditory or tactile modalities. These systems were complicated to learn but increased the accessibility of printed text. Two such examples include the *Optophone*, which used musical chords or 'motifs' [D'Albe 1914] and the *Optacon (*OPtical to TActile CONverter*)*, which used a vibro-tactile signal [Bliss 1969; Goldish and Taylor 1973]. As with our approach, the Optacon requires the user to move a camera sequentially along each line of text. However, this camera is handheld rather than worn and, more importantly, the user's other hand receives a translated vibro-tactile signal that

provides some feedback about orientation to the line. The Optacon is well researched (*e.g.*, [Hislop et al. 1983; Hislop et al. 1985]). These studies showed that even with training, reading rates (28.2 wpm on average) were limited to around 10% of the speed of sighted readers. Also, as with wearable finger-based reading approaches, Optacon users must take care in transitioning from the end of one line to the start of the next. However, even with Optacon's steep learning curve, relatively slow reading speeds, and high cognitive load, the device continues to be used by blind readers to this day, suggesting that these challenges are not necessarily a barrier to use.

With advances in sensing, computation, and OCR, modern approaches attempt to scan, recognize, and read aloud text in real-time. This transition to OCR and speech synthesis occurred first with specialized devices (*e.g.,* SARA CE*,* the original KNFB Reader[4]*,* [Gaudissart et al. 2004]), then mobile phones (*e.g.*, *Text Detective*[5], KNFB Reader iOS), and now wearables (*e.g.,* [Shilkrot et al. 2014; Shilkrot et al. 2015]*, OrCam*). While decades of OCR work exist (*e.g.,* [Mori et al. 1992; Shen and Coughlan 2012; Chen and Yuille 2004; Wang et al. 2011]), even state-of-the-art reading systems become unusable in poor lighting, require careful camera framing [Jayant et al. 2011; Manduchi and Coughlan 2014], and do not support complex documents and spatial data [Kane et al. 2013]. Because HandSight is self-illuminating and co-located with the user's touch, we expect that many of these problems can be mitigated or even eliminated.

As a wearable solution, HandSight is most related to OrCam and FingerReader [Shilkrot et al. 2014; Shilkrot et al. 2015]. OrCam is a recent commercial head-mounted camera system designed to recognize objects and read printed text in real-time. Text-to-speech is activated by a pointing gesture in the camera's field-of-view. While live demonstrations with sighted users have been impressive[6,7], there is no academic work examining OrCam's effectiveness with visually impaired users for reading tasks. A primary distinction between HandSight and OrCam is the use of hand-mounted sensing versus OrCam's head-mounted sensing. Although more work is needed to study the tradeoffs in camera position and the impact it has on function and overall user experience, HandSight's finger-based approach allows sensing of finger gestures regardless of body or head position. Secondly, HandSight supports direct-touch scanning compared to OrCam's indirect approach. Direct scanning should provide the user with increased control over reading speed as well as a better spatial understanding of a page. While our current studies focus on a finger-mounted approach, we plan to explore combining body- and head-mounted approaches in the future (*e.g.,* body-mounted for acquiring global context and finger-mounted for precise information scanning).

More closely related to HandSight is FingerReader [Shilkrot et al. 2014; Shilkrot et al. 2015], a custom finger-mounted device with vibration motors designed to read printed text by direct scanning on a word-by-word basis. As with HandSight, haptic and audio directional finger guidance is provided to help guide the reader's finger along the text. While the technical capabilities of FingerReader and the performance of its underlying algorithms have been evaluated, the two user studies reported to

[4] KNFB Reader Classic: http://www.knfbreader.com/products-classic.php

[5] Text Detective: http://blindsight.com/textdetective/

[6] OrCam TED@NYC. YouTube Video: http://youtu.be/_3XVsCsscyw

[7] OrCam at Digital-Life-Design (DLD) in Munich. YouTube Video: http://youtu.be/3m9ivtJI6iA

date have been feasibility studies with 3 and 4 participants respectively, focusing primarily on qualitative feedback. As such, it has been difficult to draw generalized conclusions about the effectiveness of finger-based reading and, more specifically, about directional guidance to support that reading. Indeed, user preference for haptic versus audio directional guidance in the two studies was conflicting. Our work addresses the need for a more in-depth investigation by comparing audio and haptic guidance with 19 participants using quantitative speed and accuracy metrics in addition to subjective feedback.

2.3. Mobile Applications for Reading Printed Text

As mentioned above, smartphone-based accessibility applications such as KNFB Reader iOS and Text Detective have begun to replace bulkier and more expensive dedicated devices like SARA CE. KNFB Reader iOS, which we consider in Study II, evolved from a dedicated scanner device in the 1970s to a Windows PDA and Nokia phone app in the early 2000s, and is now available as a $99 iPhone app [Holton 2014]. It uses the phone's built-in camera to detect and scan physical documents, then processes the captured photo and reads detected text aloud using speech synthesis (like a screen reader). The application has been well received and represents the state-of-the-art in smartphone-based document reading for the blind population [Kendrick 2008; Holton 2014]. To help users capture a document, the iPhone app provides vibration feedback based on tilt and a spoken report on document edges and rotation. Text Detective is similar to KNFB Reader iOS but has a lower user rating on the Apple App Store and did not perform as well in our own informal user tests.

Smartphones with pervasive internet access have also enabled crowd-powered accessibility approaches [Bigham et al. 2011] such as VizWiz [Bigham et al. 2010] and *BeMyEyes*[8]. Here, sighted volunteers and/or employees answer questions about visual content, including text recognition. As with mobile applications like KNFB Reader iOS, a target must be appropriately framed in the camera to receive useful feedback. Additionally, a crowdsourced approach incurs a delay for the crowd worker to respond and has negative implications for privacy. Automated applications, while perhaps not as reliable, offer more immediate results and fewer privacy concerns.

In contrast to these smartphone-based approaches, HandSight allows the user to directly touch printed text, control reading speed, and explore the spatial layout of the page. Framing may also be less problematic because the camera distance and angle are fixed to the user's finger.

2.4. Interfaces for Directional Guidance

While readers with low vision may be able to precisely trace along a line of text with their finger, blind readers will require additional finger guidance. This directional guidance needs to be responsive enough to use at reading speeds approaching braille, but should not distract the reader from listening to and understanding the text.

Haptic and audio directional guidance cues have been used in numerous projects to guide visually impaired users. Manduchi and Coughlan [Manduchi and Coughlan 2014], for example, combined synthesized speech directions with discrete audio patterns to guide users in properly framing a target for camera capture. Sonification

[8] http://www.bemyeyes.org/

has been used to assist blind users in exploring complex indoor maps [Su et al. 2010], to support non-visual learning of touchscreen gestures [Oh et al. 2013], and to convey radial direction with vowel sounds [Harada et al. 2011]. Other researchers have combined sonification and (haptic) force feedback to teach handwriting to blind children [Plimmer et al. 2011] and shapes to users with visual impairments [Crossan and Brewster 2008]. This lattermost study, for example, found that combining audio and force feedback resulted in higher user performance than force feedback alone; however, the force feedback controller moved in 3D space, so is not comparable to the finger-mounted vibration motors used in HandSight. Yet other researchers have explored primarily tactile feedback to transmit directional and shape data [Yatani and Truong 2009; Yatani et al. 2012; Noble and Martin 2006].

HandSight's audio and haptic directional guidance is designed based on these previous findings. For audio, we use pitch to convey direction, a decision based on work showing that pitch is the most salient for conveying gestural trajectory information among pitch, stereo, timbre, and volume [Oh et al. 2013]. Similar to Shilkrot *et al.* [Shilkrot et al. 2015], our haptic guidance condition varies the intensity and location of vibration on the finger to indicate direction and distance from a target (the center of a line of text).

2.5. Interfaces for Spatial Exploration

In addition to sequential text reading capabilities, blind readers may benefit from the ability to explore the layout of a printed page, locating and identifying the sizes and positions of images and blocks of text. For example, a complicated layout (*e.g.,* a two-column menu) can be difficult to interpret when scanned and read aloud by screen reader software, but by exploring the spatial positions of the text and images a user may be able to better understand the content.

Previous researchers have investigated ways to convey the spatial layout and other non-textual information on websites or touchscreens via audio or vibration. For example, Mynatt et al. [Mynatt and Weber 1994] explored methods for conveying visual user interface elements on a computer screen using braille displays and audio cues. Vibro-audio interfaces have also been explored to convey graphical information on a touchscreen such as bar graphs and visual shapes [Toennies et al. 2011; Giudice et al. 2012]. Other researchers have explored the use of sonification and tactile displays to enable blind users to access digital map data [Parente and Bishop 2003; Poppinga et al. 2011]. While our own "exploration" is simpler than these approaches (conveying the presence of text, images, and white space), future iterations can build on this body of work to enable users to more fully explore the visual content and spatial layouts of printed documents.

3. STUDY I: AUDIO VS. HAPTIC GUIDANCE FOR FINGER-BASED READING

To investigate the exploration and reading of printed text documents using finger-based interactions, we conducted a controlled lab study with 19 blind participants. The primary goal of this study was to compare audio and haptic directional finger guidance methods in terms of user performance and preference. However, as the first larger-scale study of finger-based reading (N=19 *vs.* N=3 and N=4 [Stearns et al. 2014; Shilkrot et al. 2015]), the study also quantitatively explored to what extent a finger-based reading approach can allow a blind reader to interpret the spatial layout of a document and to read and understand that document.

We simulated the experience of reading a physical document using a touchscreen tablet (an iPad) covered with a sheet of paper (Figure 2c). This approach allowed us to bypass certain technical challenges in implementing a real-time camera and text recognition system, and instead to focus on the user experience of finger-based reading. The iPad also allowed us to collect precise finger traces to enable detailed finger-movement analysis not previously possible.

3.1. Method

In this controlled lab study, participants read two types of printed documents with audio and haptic finger guidance. We used a within-subjects design with a single factor of *Directional Guidance* that had two levels (*Audio* and *Haptic*); order of presentation of the conditions was fully counterbalanced. In addition to measuring reading speed and finger movement, we collected subjective feedback and assessed basic document comprehension using standardized questions. Despite similarities to Shilkrot *et al.*'s method [Shilkrot et al. 2014; Shilkrot et al. 2015], our protocol is an extension of our previous work [Stearns et al. 2014], which was underway prior to the first FingerReader publication [Shilkrot et al. 2014]. The final apparatus and method described here were also refined through pilot sessions with 5 additional participants (1 sighted, 1 low vision, 3 blind) who did not take part in the full study.

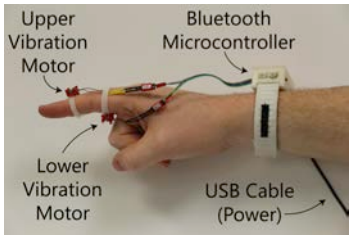
3.1.1. Participants. Twenty participants were originally recruited via campus email lists and local organizations, but one participant's data was discarded because he was unable to complete all of the required tasks. Of the remaining 19 participants, 11 were male and 8 were female, and the median age was 48 (*SD*=12.0, range 26–67). All participants were completely blind or had only minimal light perception. Five participants were congenitally blind, while the others had lost their vision later in life (some as recently as two years ago). As shown in Table 1, most participants were

ID	Age	Sex	Vision Level	Braille		Screen Reader		Computer
				Use	Comfort	Use	Comfort	Comfort
P1	54	F	Blind	5	5	4	3	4
P2	33	F	Light	4	5	5	5	5
P3	55	M	Blind	3	5	5	5	4
P4	44	M	Light	2	2	5	5	5
P5	67	M	Blind	3	4	5	5	4
P6	62	M	Light	3	4	5	5	4
P7	40	M	Blind	1	1	5	4	4
P8	27	F	Light	5	5	5	5	4
P9	49	F	Light	5	5	5	5	3
P10	43	M	Blind	5	4	1	1	3
P11	44	M	Light	4	4	1	1	1
P12	39	M	Blind	4	5	5	5	5
P13	67	M	Blind	3	3	1	1	1
P14	50	F	Light	4	4	5	5	5
P15	26	M	Blind	5	5	5	5	5
P16	48	M	Blind	5	4	5	4	4
P17	59	F	Light	2	3	1	1	1
P18	47	F	Blind	4	3	1	1	1
P19	64	F	Light	4	3	4	4	3
Mean (*SD*)	48.3 (12.0)	N/A	N/A	3.7 (1.2)	3.8 (1.1)	3.8 (1.7)	3.7 (1.7)	3.5 (1.4)

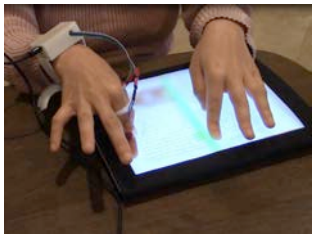
Table 1: Study I participants. All participants were either blind or had minimal light perception (denoted "Light"). Frequency of use ("Use") varied from 1 ("never") to 5 ("very often"), while comfort level ("Comfort") varied from 1 ("very uncomfortable") to 5 ("very comfortable").



(a) Screenshot of iPad software showing a single-column document.



(b) Haptic feedback device, with actuator mounted on the finger.



(c) Test setup, with physical paper covering the iPad.

Figure 2: Study I test apparatus.

frequent users of braille, although 6 were just learning to read it and rated their comfort level as lower. All but 5 participants used screen readers at least some of the time and only 4 were not comfortable with computers and/or mobile devices. Participants were compensated for time and transportation.

3.1.2. Apparatus. The test apparatus consists of an Apple iPad running custom software and connected via Bluetooth to a custom-built finger-worn haptic device (Figure 2). The source code is available on GitHub[9]. As noted previously, the iPad was used to provide a dynamic test environment that could precisely track finger movement in response to our directional guidance conditions. To simulate the feel of a physical document and reduce friction from the screen, a thin, blank paper covered the iPad. In addition, because there is no tactile border between the iPad screen and bezel, we added our own physical border made of 1/16" flexible foam (Figure 2c). The software displays documents and provides two modes of interaction: exploration and reading. All touch events (down, up, and move) on the screen are logged with x, y coordinates and timestamps.

Exploration mode. In this mode, audio cues allow users to gain a spatial sense of the document layout (*e.g.,* locations of images, columns, paragraphs) before transitioning to reading mode. As the user traces their finger over the document, they hear either a high-pitched flute sound when on a block of text or a low-pitched cello sound when on a picture. These sounds were selected and refined via pilot testing to be easily distinguishable by their pitch and timbre. When over whitespace, such as between paragraphs or columns, no sound plays.

Reading mode. In this mode, the user traces their finger from left to right along each line of text, while the system generates text-to-speech output using Apple's default iOS speech synthesis engine and provides directional finger guidance (haptic or audio depending on the condition). Reading is bimanual: the left hand, which is uninstrumented, serves as a line anchor (see "line start region" in Figure 3) while the right index finger traces the line. To begin reading, the user moves their right finger to the line start region shown in Figure 3 and an audio cue of ascending xylophone notes plays. If the finger is not already at the first line of text, audio or haptic feedback guides the user's finger up or down. Once the right finger is properly positioned over the "line start region" of the first line, the left hand joins the right hand and subsequently serves as a line anchor.

[9] https://github.com/HCIL/HandSight

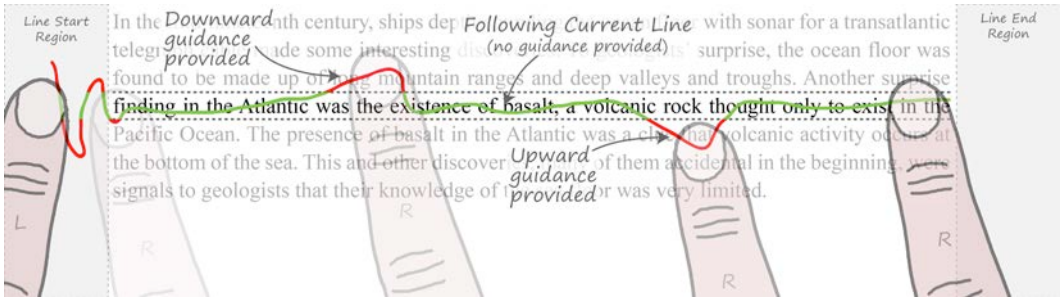


Figure 3: Reading mode interaction is bimanual. The user (1) places the right index finger in the "line start region" and moves vertically to find the start of the current line; (2) places the left index next to the right finger as an anchor; (3) traces the right finger along the line until it reaches the "line end region"; (4) returns the right index finger to beside the left finger before moving down to the next line When the right finger is directly on the line (green trace) no directional guidance is provided, but when the finger moves too high or low (red trace), audio or haptic guidance indicates which direction to move to return to the line.

The user then traces his/her finger along the line to the right, while the system speaks each word aloud and provides audio or haptic guidance whenever the finger strays above or below the line (Figure 3). The speed of the text-to-speech output adapts to match the speed of the finger movement. Speech is provided only for the current line, and only when the user's finger is within 73 pixels (0.7cm) of the middle of the line (simulating a finger-mounted camera's field of view). At the end of the line ("line end region"; Figure 3), another audio cue plays, this time with descending xylophone notes, and the text-to-speech stops. The user then moves their finger left again to find the line start region and read the next line in the same manner. Finally, at the end of a paragraph, a new audio chime plays. The audio cues for the start and end of line and end of paragraph were selected to be easily distinguishable, which we again verified using early feedback from pilot participants.

For *audio directional guidance*, the system provides a continuous tone that varies in pitch. A low pitch indicates that the finger should move down and a high pitch indicates that the finger should move up. If the finger is properly positioned over the current line, no audio plays. If the user's finger moves above the line, an audio tone at frequency 300Hz begins playing. If the user's upward movement continues, the frequency linearly decreases based on distance, down to a minimum of 200Hz at 127 pixels (1.2 cm). The 200Hz tone continues for any movement more than 2.4cm above the line. Similarly, if the user's finger strays below the line, the audio frequency begins at 500Hz and increases to a maximum of 600Hz at 127 or more pixels away. The choice to vary audio frequency to indicate distance and direction was motivated by our prior work [Stearns et al. 2014; Oh et al. 2013], and the exact pitches and distance thresholds were selected after pilot sessions.

The *haptic directional guidance* includes two disc vibration motors (8mm diameter, 3.4mm thick) controlled by an Arduino Pro Micro that communicates with the iPad via Bluetooth. The motors are attached to the user's right index finger with separate Velcro rings (Figure 4), one on top of the finger on the intermediate phalange and one below the finger on the proximal phalange. The lower motor indicates that the finger should move downward and the upper motor indicates the opposite. Neither motor vibrates while the user's finger is directly over the current line of text. Vibration intensities off the line range from a minimum perceptible

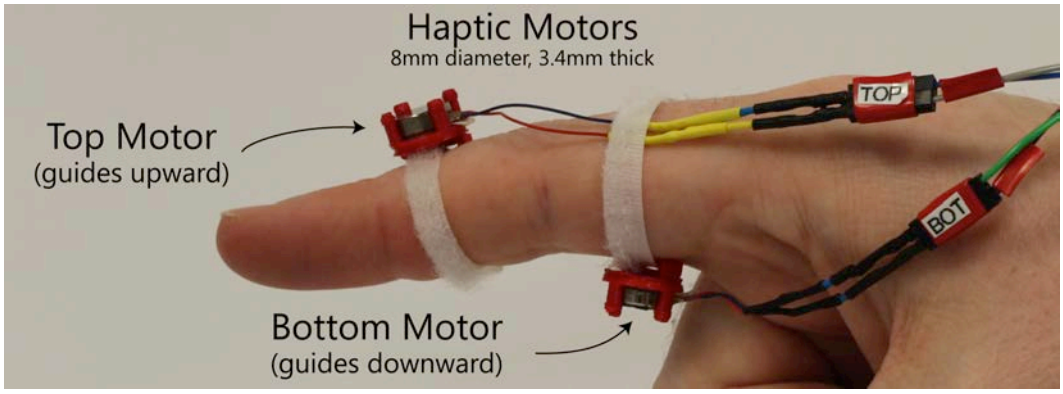


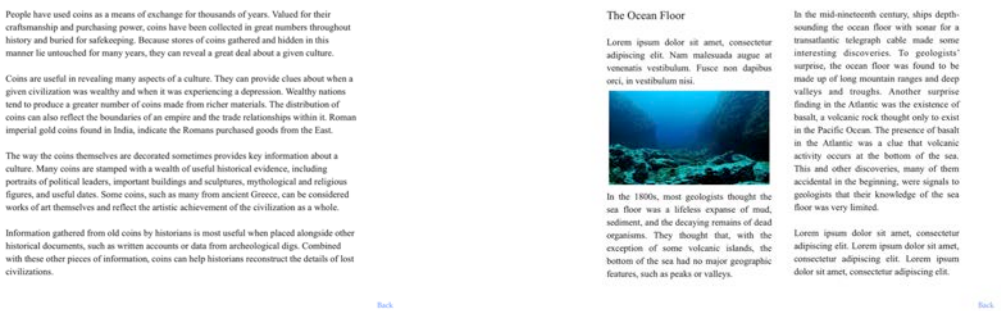
Figure 4: Close-up view of the haptic motors mounted on the finger via Velcro rings. The top motor vibrates when the user's finger moves below the line, providing upward guidance; the bottom motor vibrates when the user's finger moves above the line, providing downward guidance. The intensity of vibration depends upon the distance to the line, achieving maximum intensity at 127 pixels (~1.2 cm).

strength to the maximum strength the motors can provide, using the same distance thresholds as the audio condition. The choice to vary the position and intensity of vibration to indicate direction and distance was also motivated by our prior work [Stearns et al. 2014] and validated in pilot sessions.

In early testing within our research lab and with external pilot participants, we tested multiple mappings for audio and haptic cues and intended finger direction (*e.g.*, higher pitch to indicate up *vs.* the opposite). Users were split in terms of which mappings were most intuitive, a point we revisit in the Discussion section.

3.1.3. Procedure. Each study session lasted 1.5–2 hours. Throughout, we employed two document types (Figure 5): single-column plain text, and two-column magazine-style with a figure and an article heading. For the reading tasks described below, we adapted four test documents from a Grade 8 Iowa Test of Basic Skills practice book [Princeton Review 2000]. The original text was modified slightly for length and to ensure clarity with our speech synthesis engine (*e.g.,* removing unnecessary proper nouns); see Appendix. The documents were thus all at similar reading levels and had multiple-choice comprehension questions. We also created training documents that were similar in length to the four test documents.

Following a background questionnaire, participants first learned how to use the


People have used coins as a means of exchange for thousands of years. Valued for their craftsmanship and purchasing power, coins have been collected in great numbers throughout history and buried for safekeeping. Because stores of coins gathered and hidden in this manner lie untouched for many years, they can reveal a great deal about a given culture.

Coins are useful in revealing many aspects of a culture. They can provide clues about when a given civilization was wealthy and when it was experiencing a depression. Wealthy nations tend to produce a greater number of coins made from richer materials. The distribution of coins can also reflect the boundaries of an empire and the trade relationships within it. Roman imperial gold coins found in India, indicate the Romans purchased goods from the East.

The way the coins themselves are decorated sometimes provides key information about a culture. Many coins are stamped with a wealth of useful historical evidence, including portraits of political leaders, important buildings and sculptures, mythological and religious figures, and useful dates. Some coins, such as many from ancient Greece, can be considered works of art themselves and reflect the artistic achievement of the civilization as a whole.

Information gathered from old coins by historians is most useful when placed alongside other historical documents, such as written accounts or data from archeological digs. Combined with these other pieces of information, coins can help historians reconstruct the details of lost civilizations.

Back

The Ocean Floor

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam malesuada augue at venenatis vestibulum. Fusce non dapibus orci, in vestibulum nisi.

In the 1800s, most geologists thought the sea floor was a lifeless expanse of mud, sediment, and the decaying remains of dead organisms. They thought that, with the exception of some volcanic islands, the bottom of the sea had no major geographic features, such as peaks or valleys.

In the mid-nineteenth century, ships depth-sounding the ocean floor with sonar for a transatlantic telegraph cable made some interesting discoveries. To geologists' surprise, the ocean floor was found to be made up of long mountain ranges and deep valleys and troughs. Another surprise finding in the Atlantic was the existence of basalt, a volcanic rock thought only to exist in the Pacific Ocean. The presence of basalt in the Atlantic was a clue that volcanic activity occurs at the bottom of the sea. This and other discoveries, many of them accidental in the beginning, were signals to geologists that their knowledge of the sea floor was very limited.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Back

Figure 5: Examples of our test documents: plain text (left), and magazine (right).

document exploration mode as a precursor to the more complex task of both exploring and reading a document. The experimenter demonstrated the audio cues for text and images in exploration mode, then asked participants to explore one plain document and one magazine document for up to three minutes each. To ensure that participants understood the exploration mode, participants were asked about the structure and layout of each document (*i.e.*, how many paragraphs and columns, are there pictures or headings and if so where are they located). To avoid biasing participants toward a particular exploration strategy or interpretation, we initially provided very little direction aside from demonstrating the audio cues and warning participants of the questions they would be expected to answer. After recording the answers for a document, the experimenter then guided participants to find the correct answers to ensure that they could later use the mode correctly prior to each reading task.

After the introduction of exploration mode, participants explored and read documents with each of the directional guidance conditions (audio and haptic). The order of presentation for these conditions was fully counterbalanced. Document order was identical across all participants so that the documents themselves were matched an equal number of times with each guidance condition. To ensure similar physical experiences across conditions, participants wore the Arduino wristband and finger rings with the haptic motors throughout the study session.

The procedure for each directional guidance condition was identical, with training using a plain document (~10 minutes) followed by testing with two documents (one plain and one magazine). For the training document, the experimenter demonstrated the feedback cues and participants incrementally learned to follow a line, find the next line or paragraph, and listen to the speech feedback while moving their finger. For each test document, participants were allowed up to 90 seconds in exploration mode to assess the layout before the experimenter switched the system to reading mode. For the plain document, the reading task was to locate the first line of text and read the entire document. For the magazine document, participants read the last paragraph in the first column and the first paragraph in the second column. Exploration mode was used to locate the start of text for each document, as well as the start of the second column for the magazine document. After each test document, two multiple-choice comprehension questions provided in the Grade 8 Iowa Test of Basic Skills practice book were administered. At the end of each guidance condition, participants were asked about subjective ease of use. Finally, at the end of the study participants were asked to compare the two directional guidance conditions. See Appendix B for the full text of the subjective questionnaires.

Before conducting this study, we validated our selection of test documents and comprehension questions in a simple baseline study. Ten sighted college-age participants listened to synthesized speech of the four test documents and the comprehension questions. All 10 participants answered the questions correctly.

3.1.4. Data and Analysis. Collected data included log files from the iPad, participant responses to close- and open-form questions, and experimenter observations. To compare reading performance with haptic and audio guidance, we examined the following subtasks separately:

- *Line finding:* Finding the start of the current line. A line finding instance began with the first right-handed touch within the line start region (Figure 3)

and ended with the finger exiting that region. Sometimes participants' search paths resulted in more than one exit from the start region, so we included all data up to the final exit. For each line finding instance, we calculated elapsed time and, as an error measure, the length of the movement path traced.

- *Line tracing:* Tracing left-to-right along the current line. A line tracing instance included all touch points after a successful line finding subtask until the right index finger entered the line end region (Figure 3). For each line tracing instance, we calculated reading speed in words per minute (wpm), and, as an error measure, the average absolute distance of the finger from the vertical center of the line across all x-coordinates in that line trace.
- *Full document:* Reading the full document from the start of the first line to the end of the final line. This comprehensive analysis includes all line finding and line tracing subtasks for a single document, as well as the time to transition between columns for the magazine documents. For each document, we calculated the average reading speed in words per minute (wpm) as well as the number of skipped words that were not read aloud.

Across the 19 participants, we collected data for 1513 lines. We identified outlier samples that were more than 3 standard deviations away from the mean for a given participant and condition, removing 31 samples (2.0%) of line tracing subtask samples and 49 (3.4%) of line finding subtask samples.

We used paired t-tests to compare line tracing speed between haptic and audio guidance. However, other measures violated the normality assumption of a t-test (determined using separate Shapiro-Wilk tests for each measure, $p < 0.05$). For these measures, we conducted non-parametric Wilcoxon signed rank tests to compare haptic and audio. For all posthoc pairwise comparisons, we applied Holm's sequential Bonferroni adjustments to protect against Type I error [Holm 1979].

3.2. Findings

Our findings include performance results derived from the log data and exploratory descriptions of how participants responded to and interacted with the finger-based reading approach (*e.g.*, initial use of exploration mode, potential advantages of such an approach).

3.2.1. Reading Mode – Line Tracing. Figure 6 shows line tracing performance. For plain documents, the average reading speed with haptic guidance was 120.9 wpm (*SD*=57.0), compared to only 106.3 wpm (*SD*=46.2) with audio; however, a paired t-test comparing the two types of guidance was not statistically significant. A similar trend followed for magazine documents, at 111.8 wpm (*SD*=43.3) and 106.7 wpm (*SD*=54.1) for haptic and audio, respectively, with a paired t-test revealing no statistically significant difference between the two.

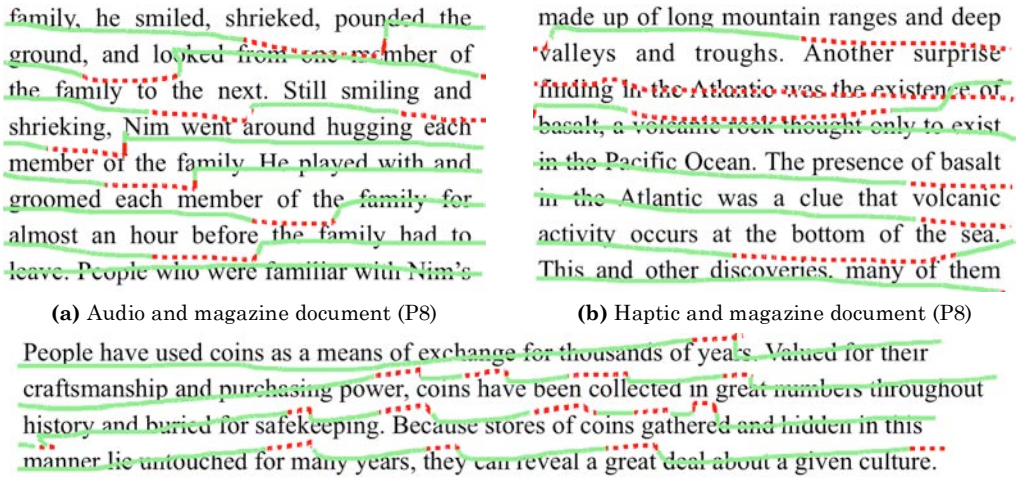
In terms of error, audio guidance was significantly more accurate than haptic guidance for the magazine documents, with an average distance of 11.2 px (*SD*=3.5) to the center of the line versus 14.6 px (*SD*=5.7). A Wilcoxon signed rank test was statistically significant on this measure, with a large effect size (Z_{19}=-2.374, p=.018, r=.54). Figure 7 shows a representative finger trace that illustrates this performance difference. For the plain documents, however, the two guidance conditions resulted in more similar distances, at 11.9 pixels for audio (*SD*=4.6) and 12.8 pixels for haptic (*SD*=4.6). This difference was not significant using a Wilcoxon signed rank test.

Participants tended to drift frequently, spending on average 29.7% (*SD*=13.2) of their line tracing time off of the line for the audio condition and 37.8% (*SD*=14.7) for the haptic condition. Reflecting the distance accuracy results above, this difference was statistically significant with a Wilcoxon signed rank test (Z_{19}=-2.57, p=.010, r=.59). In addition, participants tended to drift consistently above or below the line. Figures 7a, for example, illustrates downward drift whereas Figure 7c shows upward drift. We observed 11 participants who drifted consistently upward, 4 who drifted consistently down, and 4 who varied by document or did not tend toward either direction. This tendency may have been affected by how each participant's arm was positioned relative to the iPad—participants were instructed to rotate the screen as needed, but few chose to do so.



(a) Average line tracing speed **(b)** Average line tracing error

Figure 6: Average line tracing speed (higher is better), and average error—vertical distance offset from the center of the line (lower is better). Error bars indicate standard error (*N*=19). Performance was generally similar between the audio and haptic conditions, but audio resulted in significantly lower line tracing error for the magazine document (*).



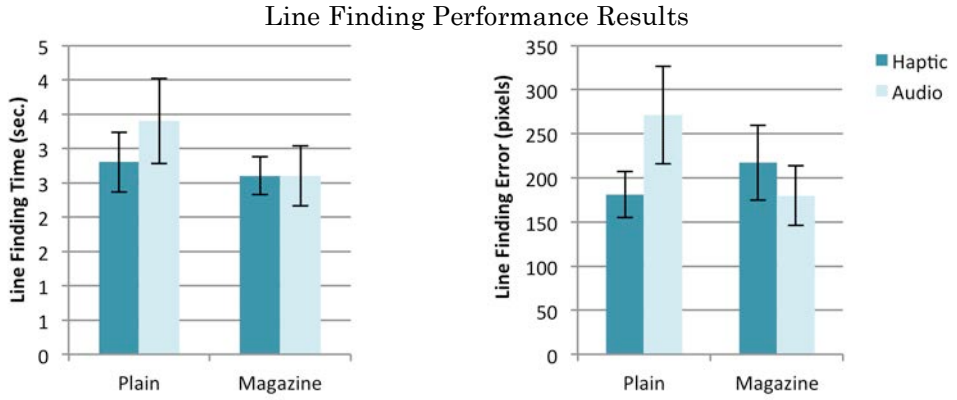
(a) Audio and magazine document (P8) **(b)** Haptic and magazine document (P8)

(c) Audio and plain document (P7)

Figure 7: Example finger traces. Solid (green) indicates that the finger was on the line, while dotted (red) indicates that the finger was off the line and directional guidance was being provided. (a) and (b) illustrate the difference in accuracy between the audio and haptic guidance conditions for P8. Participants frequently reacted more immediately to audio guidance, but tended to ignore small amounts of vibration with haptic guidance. This observation may explain the significant difference in error between the audio and haptic conditions. Participants also tended to drift consistently above or below a line as they read, as seen in (a), (b) and (c).

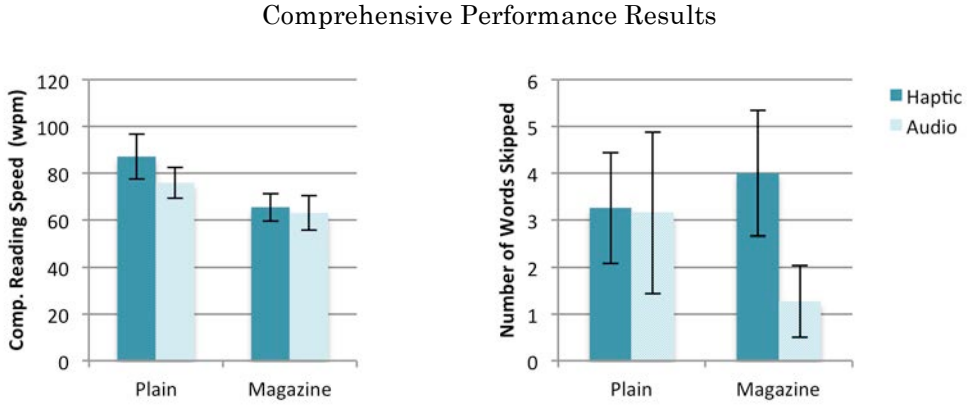
3.2.2. Reading Mode — Line Finding. As shown in Figure 8, line finding performance was similar across all directional guidance conditions and document types. No significant differences were found between haptic and audio guidance for either document type or performance measure using Wilcoxon signed-rank tests. Across all conditions, it took participants on average 2.6–3.4 seconds to find the next line in a document (plain: haptic M=2.8 seconds, SD=1.9, and audio M=3.4, SD=2.7; magazine: haptic M=2.6, SD=1.7, and audio M=2.6, SD=1.9). For error, measured as the average path length while searching for the start of a line, haptic averaged 181.2 pixels (SD=114.0) with plain documents, while audio averaged 270.9 pixels (SD=241.1). In contrast, for magazine documents, haptic averaged 217.1 pixels (SD=184.4), compared to 179.8 pixels for audio (SD=147.5). Again, however, these differences were not found to be statistically significant.

3.2.3. Reading Mode — Overall Performance and Comprehension. Figure 9 shows comprehensive performance over the documents, including the total reading time and



(a) Average line finding speed (b) Average line finding error

Figure 8: The average time elapsed (left) and error (right) in finding the next line; lower is better for both graphs. The error bars indicate standard error (N=19). Performance differences between the two conditions were not significant.



(a) Average comprehensive reading speed (a) Average number of skipped words

Figure 9: The comprehensive reading speed for an entire document (higher is better) and total number of skipped words (lower is better) by document. The error bars indicate standard error (N=19). Performance differences between the two conditions were not significant.

number of skipped words. Reading speeds ranged from 63–81 wpm (plain: haptic *M*=81.1 wpm, *SD*=42.1 and audio *M*=75.8 wpm, *SD*=29.0; magazine: haptic *M*=65.4, *SD*=25.6 and audio *M*=63.0, *SD*=31.9). Overall, the number of skipped words, that is, words that were not read aloud by the text-to-speech engine, was uniformly low across conditions. The four documents contained an average of 211.5 words, but only 1–5 of those words were skipped on average for any given document. The number of skipped words was also similar between conditions for the plain documents (plain: haptic *M*=3.3, *SD*=5.2 and audio *M*=3.2, *SD*=7.5; magazine: haptic *M*=4.0, *SD*=5.9 vs. audio *M*=1.3, *SD*=3.3). Using Wilcoxon signed-ranks tests, no significant differences were found between haptic and audio guidance for either measure (speed, number of skipped words) with either document type.

While further investigation is needed to determine to what extent audio and haptic guidance impact comprehension, overall, participants answered the comprehension questions with high accuracy. Across all participants and conditions, 85% of the questions were answered correctly (Table 2).

3.2.4. Overall Subjective Response. Overall preference was split, with a small majority of participants (11 out of 19) preferring haptic feedback, 7 preferring audio, and 1 reporting equal preference. Participants also rated the two types of guidance in terms of comprehension and line tracing ease, from *1 – very difficult* to *5 – very easy*. The ratings, shown in Table 3a, support the overall preference patterns. Both guidance conditions were rated somewhat positively for both measures (3.1 or higher on average), and the differences between the two conditions were not statistically significant with Wilcoxon signed rank tests. Some challenges with the HandSight approach were seen as common to both types of guidance. For example, P12 said,

Guidance	Document	2/2 Correct	1/2 Correct	0/2 Correct
Audio	Plain	14 participants	3 participants	2 participants
Haptic	Plain	17	2	0
Audio	Magazine	12	5	2
Haptic	Magazine	14	5	0

Table 2: Number of participants who answered the set of two comprehension questions correctly in each experimental condition (*N*=19). Most questions were answered correctly regardless of condition.

		Question	*N*	Mean	*SD*
(a)	Ease of use: Haptic vs. audio	Reading comprehension with audio guidance	19	3.2	1.3
		Reading comprehension with haptic guidance	19	3.7	1.2
		Line tracing with audio guidance	19	3.3	1.3
		Line tracing with haptic guidance	19	3.1	1.4
(b)	Ease of use: Elements common to both conditions	Start of text detection	19	4.5	0.7
		Start of line detection	19	4.2	0.6
		End of line detection	19	4.7	0.5
		End of paragraph detection	19	4.6	0.8
		Start of column detection	19	3.8	1.2
(c)	Comparison to existing technologies	HandSight vs. braille	18	3.0	1.0
		HandSight vs. screen readers	14	2.9	1.2
		HandSight vs. other reading aids	12	2.4	1.2

Table 3: Study I subjective ratings from 1 to 5 where 5 is the best. **(a)** Reading comprehension and line tracing for each guidance condition. **(b)** Experience with subtasks common to both guidance conditions. **(c)** Overall comparison (better/worse) of HandSight versus braille, screen readers, and other reading aids. A score of 5 indicates that HandSight was perceived as much better than the existing technology, while a score of 1 indicates that it was much worse.

"The haptic feedback only tells you when you're not in line, not where the next thing would be", and made a similar comment for audio guidance.

The 11 participants who preferred haptic guidance generally felt that it was more intuitive, easier to use or faster than the audio. For example, P13 stated: *"It gave me a clearer indication of which way, up or down"*. P9 also commented, *"The vibrations kind of helped as a prompt, so that I automatically would go in the right direction, and I was able to read faster"*. Six of the participants who preferred haptic guidance also mentioned that the audio guidance was more distracting, and that made it harder to focus on the speech feedback: *"You could focus on the audio of the text, and not be listening for other sounds"* (P7), or *"I missed a couple words because I was being distracted by the [audio]"* (P15). Even 4 of those who preferred audio guidance mentioned that the overlapping sounds could be somewhat distracting.

Of the 7 participants who preferred audio, almost all (N=6) found haptic guidance to be confusing: *"Sometimes when I use the vibrations I would forget which direction I was going based on where the vibration was"* (P5), or *"I had to analyze more what the vibrations meant"* (P14). Two participants also mentioned concerns about comfort, especially for prolonged use, for example: *"If you're reading longer your finger might get numb and it might get more difficult to figure out where the vibration was"* (P14).

Participants found the audio cues common to both guidance conditions relatively easy to use. Using these audio cues to detect the start of the text, line start/end areas, start of a column, and end of paragraph were all rated above 3.8 on a 5-point scale (Table 3b). Detecting the start of a column received the lowest score (M=3.8), perhaps reflecting the challenge of reading text with a more complex layout. This challenge can be non-trivial for some users. It should be noted that the participant whose data we discarded (described in Section 3.1.1), had been blind since early childhood and was thus unfamiliar with the concept of a two-column document, an issue that requires further consideration in future work. He said: *"Can a document be structured this way, with a paragraph just taking half part of the page?"* Other participants also found the magazine document to be more difficult, especially those who were congenitally blind, but all were able to successfully complete the task.

3.2.5. Comparison to Other Technologies. As shown in Table 3c, the overall experience of HandSight was rated similarly compared to braille (M=3.0, SD=1.0), and somewhat negatively compared to other aids such as cell phone apps or scanner hardware (M=2.9, SD=1.2), and screen reader software (M=2.4, SD = 1.2).

Seven participants who were not comfortable with braille or existing reading technologies generally liked the lower learning curve and flexibility of our reading approach. For example, P11, who was currently learning braille said: *"With braille you gotta always constantly remember which dots are for which letters […]. this will tell you what the word is. Less stress."* (P11). P7 also commented on the utility of being able to directly control reading speed with our approach: *"A [screen] reader you get like one speed, it doesn't slow down for any reason, and sometimes it's a lot harder to go back and get your place from where you stopped."*

However, nine participants who were more familiar with braille and other reading devices raised concerns about ease of use and cognitive load. P14, for example, preferred braille: *"Reading braille I can read at a steadier pace and I can know where the punctuation is, and it's easier for me to find the next line"* (P14). Both P16 and P18 commented on cognitive load: *"There's the need to concentrate on staying within lines"*

(P16), and, *"I'm so focused on trying to read the document, I'm not necessarily retaining the information the way I want to"* (P18).

3.2.6. Initial Use of Exploration Mode. The analyses above focus on reading mode, but at the start of the study, participants first used exploration mode to receive feedback on the presence of text, images and whitespace in both plain text and magazine documents. Even with this initial use, all but one participant correctly identified the presence or absence of a picture in both documents and described the picture's location. Determining whether audio breaks represented a gap between two paragraphs or two columns was more difficult, such that 11 participants initially identified multiple columns in the plain text document. However, between the two documents, the experimenter revisited how to distinguish between paragraphs and columns, and almost all participants (17 out of 19) were able to report the correct number of columns for the magazine document. Precisely counting paragraphs was still difficult, with only 9 and 3 participants reporting the correct number for the plain and magazine documents, respectively. For the magazine document the primary source of error was confusion over the definition of a paragraph in a multi-column document—the majority (N=15) did not count the paragraphs in the two columns separately. Additionally, 7 participants mistook the heading in the magazine document for another paragraph, and only 9 answered questions about it correctly.

We observed a few exploration strategies, with some participants using multiple strategies: 8 initially moved their fingers quickly but in no discernible pattern, searching out the locations of images and text within the document; 8 followed a procedure similar to reading braille, exploring left to right sequentially down the page; 12 explored sequentially left to right then top to bottom, counting breaks in the sound to identify paragraphs and columns. Though we only told participants that we would ask them about the number of paragraphs, columns, and the presence/location of certain features (*e.g.,* headings, pictures), 6 participants provided additional details such as the width of the margins and the size and locations of the images and blocks of text.

Four participants provided unprompted feedback that they liked the document layout knowledge provided by the exploration mode. P6, for example, compared this advantage of the finger-based approach to a traditional screen reader:

> *"You have a perspective of the document layout—how many columns, where the graphics are located, the heading, and things like distribution of the text itself. […] When you use screen readers, you don't have any idea about that, you just get the text, you just get the content, but you don't have any direct access or idea of the document layout or things like that."* (P6)

P15 was particularly excited about the idea, using the exploration mode to identify the size and locations of images and blocks of text, and speculating based on their relative positions that *"maybe [this block of text is] a description of the picture. I always wonder things like that."* In contrast, P12 stated that he didn't see a use for spatial information in most situations: *"Not for blocks of text, but […] for diagrams or for maps it might be, because that's the only time that you actually need spatial orientation on a page."* He felt that a system that could automatically process a page and abstract the layout would be preferable. Further investigation is needed to evaluate how much this additional spatial information impacts comprehension or

document understanding, as well as how to best present that information to the user via audio or haptic feedback. We return to this point in the Discussion section.

3.2.7. Summary of Study I Findings. Audio and haptic guidance resulted in relatively similar user performance, although audio may offer an accuracy advantage for line tracing with some documents (it was significantly better than haptic for the magazine document). Although the majority of participants preferred haptic guidance, the overall split in preference reflects contradictions found in previous research [Stearns et al. 2014; Shilkrot et al. 2014; Shilkrot et al. 2015]. Open-ended comments also highlight the tradeoffs of the two types of guidance, such as the interference of audio guidance with speech output and the potential for desensitization with haptic guidance. Finally, while several participants appreciated the direct access to layout information provided with HandSight's exploration mode, and the lower learning curve of HandSight compared to braille, important concerns arose about ease of use and the amount of concentration required.

4. STUDY II: PRELIMINARY USE OF A PROOF-OF-CONCEPT PROTOTYPE

Following the in-depth comparison of audio and haptic finger guidance in Study I, we recruited 4 participants to return and provide qualitative feedback on a proof-of-concept wearable prototype. These follow-up sessions allowed us to collect preliminary evidence of: (1) the extent to which a blind reader can use a finger-mounted camera and directional guidance system to explore and read a printed document, and (2) the strengths and weaknesses of finger-reading versus a mobile scanner and screen reader.

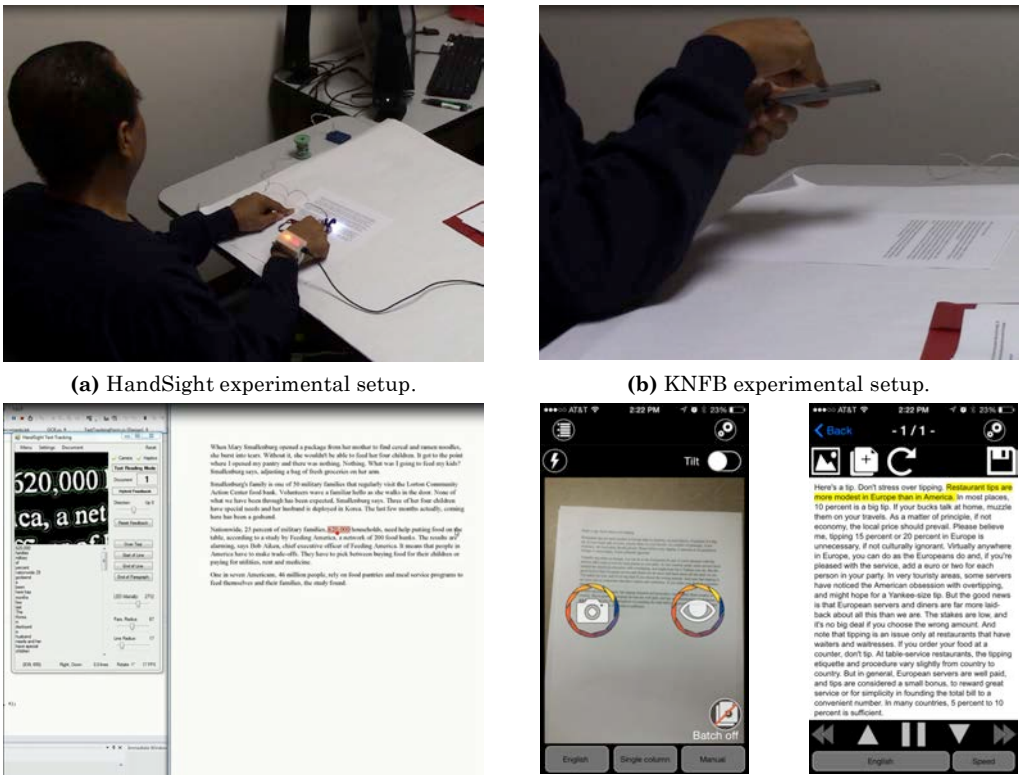
4.1. Method

Participants explored and read printed documents using a proof-of-concept finger-mounted camera system, followed by KNFB Reader iOS, a popular mobile document reader. This was not intended to be a controlled comparison of the two technologies, but instead allowed for preliminary user experience feedback.

4.1.1. Participants. We randomly selected 4 participants from Study I to return for this follow-up study, with the constraint that they represent a mix of preferences for haptic and audio directional guidance. Study II was conducted shortly after Study I was completed, with participants returning between 1 and 3 weeks after their initial session. Participants' durations of blindness varied from 2 to 30 years, but none were congenitally blind. Only one participant (P12) had experience with KNFB Reader iOS. Refer to Table 1 for demographic information and to Table 4 for experience with specific technologies, including KNFB Reader iOS. As with Study I, participants were compensated for their time and transportation costs.

ID	Study 1 Feedback Preference	Frequency of Braille Use	Frequency of Screen Reader Use	Familiar with KNFB Reader iOS?
P10	No Preference (Tested Audio)	5	1	No
P11	Haptic	4	1	No
P12	Audio	4	5	Yes
P19	Haptic	4	4	No

Table 4: Study II participants; IDs are carried over from Study I. Comfort levels ranged from 1-5, with 1 indicating "very uncomfortable" and 5 indicating "very comfortable".



(a) HandSight experimental setup.

(b) KNFB experimental setup.

(c) Screenshot of HandSight software.

(d) Screenshots of KNFB Reader iOS.

Figure 10: Study II experimental setup. (a) The HandSight test apparatus consisted of a desktop computer running a custom reading program, stereo speakers, a finger-mounted camera system, and the haptic feedback device from our first study. Participants were asked to read through two documents using our prototype system. (b) The KNFB experimental setup consisted simply of an iPhone with the KNFB Reader iOS app. Participants were asked to read three documents using the app. (c) A screenshot of HandSight's OCR interface (this was not shown to the participant and used only by the experimenter). (d) Two screenshots of KNFB Reader iOS: (left) the 'capture' interface helps users orient the phone's camera to take a photo of the target document; (right) the digitized document screen-reading interface.

4.1.2. Apparatus. The proof-of-concept HandSight prototype consisted of a desktop computer running custom software, external speakers, a finger-mounted camera, and the haptic device from Study I (Figure 10). The camera was a self-illuminated Awaiba NanEye 2C CMOS camera and LED ring (~40 fps, 90° square field of view, 250x250 pixels, 2.4mm diameter), embedded in an adjustable ring and positioned above the finger to point down at the page (Figure 1b). The camera was positioned 1–2cm above the page, and had a field of view approximately 1.5cm across (2–3 lines of text). These numbers varied somewhat depending on the participant's hand position.

As with Study I, the software provided two modes of interaction: *exploration* and *reading*. Exploration mode provided the same feedback as in Study I, except that the prototype system did not detect images; as such, documents used in Study II did not include images. To ensure that the flute sound did not stop between individual characters or lines of text, the system first blurred the text using a blur radius that was manually calibrated prior to beginning the exploration tasks. The audio and

haptic cues in reading mode were identical to those in Study I, with text-to-speech output using the IVONA Voice for Windows speech synthesis engine.[10] Exploration and reading events were logged with timestamps, but we could not log precise finger-trace data as we had done with the iPad in Study I.

The software processed each video frame from the camera using OpenCV[11], an open-source computer vision and image-processing library. With each frame, we applied four preprocessing algorithms. First, to correct radial distortion from the camera lens, we used standard camera calibration algorithms [Hartley and Zisserman 2003]. Second, to reduce noise, perform binarization necessary for OCR, and adapt to uneven lighting from the LED, we filtered each frame using an adaptive threshold in a sliding window. Third, to reduce false positives, we performed a connected component analysis and removed components with areas too small or aspect ratios too narrow to be characters. Finally, to correct for finger rotation, we blurred the image to efficiently group the components into likely lines of text, then extract the minimum-area bounding rectangle for each new component. We used the estimated orientation of this rectangle to correct for camera rotation, inverting it so that the lines of text were parallel to the x-axis. This process is similar to that described in [Safabakhsh and Khadivi 2000].

To simplify sensing for this proof-of-concept prototype, we assumed that a complete image of the page was available to the system in advance. The software then estimated the current finger location by performing OCR on the visible text and matching it to the known content of the page. We used the Tesseract OCR library[12] for text detection and recognition of each preprocessed frame, then compared the results to the pre-computed document text. For efficiency, we tracked character motion between frames and only performed OCR when sufficient motion had occurred or when the system was unable to reliably estimate the current location (allowing us to achieve an average processing rate of 20–30 fps). Because the camera's field of view was large enough to encompass multiple partial words across 2-3 lines of text (Figure 10c), the system did not generally encounter difficulty distinguishing the locations of repeated words. The likelihood of this potential problem was further reduced using recent location estimates and the motion of the user's finger to resolve conflicts. We tracked the current line of text using the camera's estimated motion and the known content of the page, and only provided text-to-speech feedback when the user advanced on the current line. In order to provide a smooth reading experience, it was not possible to skip or repeat words. Although this enforced sequential reading of the text, it mitigated several potential sources of confusion that would have arisen had we allowed rereading or moving between lines. The software detected that the user had reached the start or end of a line or paragraph using the known content of the page, and provided the same audio cues as in Study I. Also, as with Study I, the speed of the text-to-speech feedback was adjusted to match the user's finger speed.

The test apparatus for the second part of the study consisted of the KNFB Reader iOS application running on an iPhone 5S with the VoiceOver feature enabled. To take a picture, users tapped on the left side of the screen to select the "Take Picture"

[10] http://www.ivona.com/us/for-individuals/voices-for-windows/

[11] http://opencv.org/

[12] https://code.google.com/p/tesseract-ocr/

button, and then double-tapped the button to capture an image. The software played a shutter sound to inform the user that the picture was captured successfully, and then immediately began reading any recognized text.

4.1.3. Procedure. These exploratory study sessions lasted 1–2 hours. The participant first used HandSight with his or her preferred directional guidance method from Study I. As with Study 1, training and testing documents were selected from the *Iowa Test of Basic Skills*. For training, the experimenter first re-introduced exploration mode and asked the participant to explore a plain document for up to three minutes. Participants were directed to count the number of paragraphs and columns, and to note the size and position of the margins. The experimenter then re-introduced reading mode's audio cues and directional guidance, and helped the participant read the training document, providing verbal or physical guidance if necessary. The training tasks lasted 10–15 minutes. After training, participants explored and read one single-column test document, with the experimenter providing verbal assistance only if the participant was unable to proceed. Afterward, participants answered questions about the layout of the document, three multiple-choice questions to judge comprehension, and subjective questions about the experience. We did not use a magazine-style document because HandSight does not currently support two-column documents.

Following the use of HandSight, the experimenter introduced KNFB Reader iOS: how to position the phone's camera over a page, take a picture, and listen to the recognized text. Although the KNFB Reader iOS application included a spoken field of view report to assist with framing a document, we did not evaluate this feature due to time constraints and because it was not the focus of this study. Participants were allowed to repeat this process up to three times with a single-column training document, with verbal or physical guidance as needed. This training task lasted 10–15 minutes. Participants then read two test documents unassisted: a single-column document (from the Iowa Test of Basic Skills) and a two-column magazine document (from USA Today) similar to those read in Study I but without images. KNFB Reader iOS advertises support for multicolumn formats, and the procedure for capturing and reading the two types of document was identical. If the participant was unsatisfied with the reading result, they were allowed one additional attempt per document. Participants answered multiple-choice comprehension questions after the single-column document and summarized the content of the two-column document. Finally, participants reported on their experience using the application. See Appendix B for the full text of the subjective questionnaires for both HandSight and KNFB Reader iOS.

4.2. Findings

Our findings are exploratory, including general observations about how participants approached the reading tasks, and subjective responses to both our proof-of-concept implementation and KNFB Reader iOS. While the focus is on qualitatively describing experiences with the technologies, we include performance statistics such as reading speed, line finding time, and number of skipped words.

4.2.1. Overall Experience. All four participants completed the reading tasks, but with varying levels of success (Table 5). P10 read slowly and required frequent verbal and physical intervention by the tester to adjust hand position and answer questions

Participant Identifier	P10	P11	P12	P19	Mean
Guidance Type	Audio	Haptic	Audio	Haptic	N/A
Start of Text	5	2	5	1	3.3
Start of Line	5	2	2	2	2.8
End of Line	5	5	5	5	5.0
End of Paragraph	3	5	5	5	4.5
Line Tracing	2	2	3	2	2.3
Understanding Cues	5	5	5	3	4.5
Reading and Understanding	3	3	5	4	3.8
Mean Ease of Use Rating	4.0	3.4	4.3	3.1	3.7
Average Reading Speed per Line (wpm)	18.4 (*SD*=5.5)	56.6 (*SD*=16.4)	60.2 (*SD*=11.1)	44.9 (*SD*=17.1)	45.0
Average Line Finding Time per Line (s)	30.5 (*SD*=24.4)	8.8 (*SD*=5.6)	7.3 (*SD*=5.0)	18.0 (*SD*=12.7)	16.15
Time to Read Full Document (s)	1493	469	409	717	772
Comprehension Questions Score	2/3	3/3	3/3	3/3	2.75/3

Table 5: Top: Ease of use responses while using the HandSight prototype. Responses range from *1 - very difficult* to *5 - very easy*. Bottom: Performance metrics from the HandSight reading task. The document for this task consisted of 282.6 words (normalized to 5 character length) across 17 lines.

about the directional audio cues. P11 and P19 read more quickly, and only needed infrequent verbal reminders (P11 was reminded once about hand position, and P19 was reminded once about hand position and the procedure for finding the start of a line). P12, who was very comfortable with both braille and screen readers, read the fastest, at 60.2 wpm, and did not require any assistance. Only P10 failed to answer all three comprehension questions correctly, likely due to decreased attention to the content while struggling to complete the task.

Comments were similarly mixed. P19 was enthusiastic about the concept, stating: *"I'm very pleased and excited about the system. I think it could make a great difference in my life."* P12 was more critical, finding the approach to be slower than expected: *"It seems like a lot of effort for reading text."* P10, P11, and P19 were all learning to read braille at the time of the study, and P11 and P19 found the reading experience using HandSight to be easier than braille for reasons similar to those expressed in Study I (*e.g.,* lower learning curve, less to remember). P10 stated that braille and finger-reading were both difficult at times, requiring too much concentration to read quickly or fully comprehend the text. P12, who had the most braille experience, found HandSight to be "much worse" than braille and "somewhat worse" than other technologies for reading printed documents. In addition to commenting on the ease of following a line of braille text due to the tactile feel of the dots and the lack of layout issues such as multiple columns, P12 said that he typically scans printed documents to read on his computer or mobile device, an approach he finds faster compared to HandSight and one that does not require use of both hands.

4.2.2. Cognitive Load. Although they were able to complete the reading task, all participants expressed concern about the level of concentration required to interpret the directional guidance and other audio cues while listening to synthesized speech. P11, for example, commented on the difficulty of remembering how to map the haptic guidance to up/down movement: *"it gets you a little confused sometimes, especially if you was [sic] into reading the story and you forget which one was the vibration for moving up and which one was for moving to the bottom."* P11 also commented on the focus and practice required, concluding that it would be difficult to use, *"if you're tired, if you've had a long day."* More practice with the device may address some of



(a) (b) (c) (d) (e)

Figure 11: Examples of situations where HandSight was unable to provide feedback. All images have been preprocessed to emphasize text and highlight baselines for the current line. (a), (b) Not enough text is visible in the margins to provide directional guidance. (c) The camera position changed after calibration, and too far from the page to reliably recognize text. (d) The camera is moving too quickly, blurring the text and reducing the frame rate of the recognition algorithms. (e) The user's middle finger is in the camera's field of view, preventing correct segmentation of the lines of text.

these issues, though interaction design changes are also likely needed (e.g., more intuitive and responsive directional cues to reduce required concentration on line tracing task, efficient rereading to enhance comprehension).

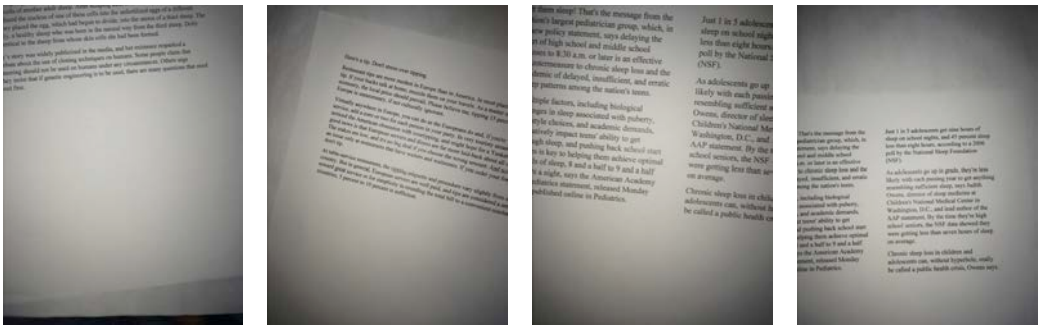
Technical limitations with the prototype may have exacerbated cognitive load issues. Although our algorithms ran at approximately 30 fps on average, they tended to run more slowly after rapid finger movements. This limitation caused a noticeable lag at times, which P11 and P19 reported required more concentration. P19, for example, commented that after the start-of-line audio cue there was sometimes a delay before the speech began, causing problems: "*I wasn't getting that in my head to just wait for the delay. I started moving my finger*".

4.2.3. Physical Design. Three participants identified limitations with the prototype's physical design. The primary issue stemmed from the camera placement: for the text to be an appropriate size and orientation within the camera's field of view, participants' hands needed to be held at a specific angle. Although the camera's placement on the finger was adjusted at the start of each study session, it could not easily be readjusted. Participants thus had to hold their hand at nearly the same angle throughout the study. Two participants reported that this position was too uncomfortable for extended use, suggesting that the physical design will need to improve in future versions and/or the camera location should be easily adjustable. We also identified the need for feedback when the system loses its position in the text or is unable to recognize visible text in reading mode (Figure 11 shows examples); this occurred when the hand position changed too much or, more commonly, when the participant moved into the upper or lower margins of the document.

4.2.4. HandSight vs. KNFB Reader iOS. While the study did not offer a controlled comparison of HandSight and KNFB Reader iOS, we can draw preliminary conclusions about tradeoffs between the two. Even without KNFB Reader iOS's document-framing guidance enabled, participants unanimously preferred it to HandSight, with three participants rating it as *5 – much better* and one as *4 – somewhat better*. The primary reason was the fluidity of the reading experience after capturing an image with KNFB—the application read the full document quickly and participants were able to concentrate solely on the content of the passage. For example: *"It just did it all for you, that way you just listen to what it's saying and then take in the details"* (P11). P12, who had previous experience with KNFB, also stated:

Participant Identifier	P10	P11	P12	P19	Mean
Document 1: Number of Attempts	2	2	1	2	N/A
Document 2: Number of Attempts	2	2	1	2	N/A
Document 1: Total Time (s)	230	198	93	225	187
Document 2: Total Time (s)	138	219	89	137	146
Document 1: Text Lost (%)	29.7%	48.6%	0.4%	10.4%	22.3%
Document 2: Text Lost (%)	51.5%	0.0%	0.0%	51.5%	25.8%
Comprehension Questions Score	1/3	3/3	3/3	3/3	2.5/3

Table 6: Performance metrics from the KNFB Reader iOS reading tasks. The amount of text lost includes both cropped and misrecognized words, and the percentages indicate the best performance out of the two attempts participants were allowed for each document.



(a) **(b)** **(c)** **(d)**

Figure 12: Examples of cases where the KNFB Reader iOS application failed to fully capture the content of a document due to partial visibility or excessive rotation.

"I like that the text is immediately available to use for other purposes […] I can go back and review the text letter by letter if I need to." The average reading time was only 187 seconds for the first document and 146 seconds for the second, even with the two attempts that participants were allowed, as compared to an average time of 772 seconds to complete the reading task with HandSight (Tables 5 and 6).

Although participants preferred KNFB Reader iOS overall, the process of capturing an image was not always straightforward without the document-framing guidance. P11, for example, said: *"It was easy to read it once you got it right, but it was difficult to center [the camera] in order to get the whole text"* (P11); see Figure 12 for examples of images captured by participants during this study. With the exception of P12, all participants required a second attempt to capture each document, and even with a second attempt part of the document was frequently omitted. Although accuracy varied across participants and attempts, approximately one quarter of the documents' content was missed on average (see Table 6).

Document comprehension appeared to be similar to reading with HandSight, with three participants answering all comprehension questions correctly (Tables 5 and 6). However, even when participants were able to understand the main points of a document, the reading experience was not always smooth due to missing text: *"It's not always easy to know if I have the entire page. That was a problem with the first test document. While it was still understandable, I clearly lost some of the text"* (P12).

4.2.5. Summary of Study II Findings. While this study was not meant to be a controlled comparison of finger-based reading versus a mobile scanner, it offers some guidance for future studies. HandSight provided more immediate access to text content than KNFB Reader iOS, but was much slower and was perceived as requiring a greater level of concentration. Once the document was successfully scanned, KNFB Reader iOS offered a faster and smoother reading experience, and was preferred by all participants. HandSight provided additional information about the spatial layout of documents, but further investigation is needed to determine the impact that may have on document understanding.

5. DISCUSSION

Our findings highlight tradeoffs between haptic and audio directional guidance for finger-based reading. We also reflect on the feasibility of finger-based reading compared to existing methods, and outline ideas for iteration on HandSight's design.

5.1. Audio versus Haptic Directional Guidance

For blind users, effective finger guidance is critical for line-by-line reading, and therefore directly impacts the feasibility of the finger-based reading approach. Although there were few statistically significant differences between audio and haptic finger guidance in Study I, some tradeoffs emerged. For the magazine documents, audio guidance resulted in significantly more accurate line tracing than haptic guidance. The exact cause is unclear. To scaffold participants in learning how to do finger-based reading, we always presented plain documents before the more complex magazine documents. That audio was more accurate than haptic for the magazine document thus suggests that haptic may have a steeper learning curve, participants have become desensitized to the haptic vibration over time, or that haptic is somehow not as effective with complex document layouts.

In terms of subjective responses, our findings reflect the conflicting results seen in prior work [Stearns et al. 2014; Shilkrot et al. 2014; Shilkrot et al. 2015]. Out of 19 participants, 11 preferred haptic, 7 preferred audio, and one was undecided. One downside of the audio guidance is that it occupies the same channel as the speech output, which made it difficult for some participants to concentrate on the text-to-speech synthesis. Twelve participants, half of whom even preferred audio guidance, commented on this issue. For haptic guidance, the potential issue of desensitization or numbness arose even in this short study, suggesting that a longer-term evaluation will be important.

We also encountered disagreement over how audio and haptic cues should map to up/down direction, which could have impacted results. The mappings used in our studies were the result of pilot testing and our experiences in [Stearns et al. 2014]. For audio, we used high pitch to indicate that the finger should move up and low pitch for down. For haptic, the vibration motor on the underside of the finger indicated downward movement, while the top vibration motor indicated upward movement (in essence, pulling the finger). While the majority of Study I participants were satisfied with these mappings, 4 felt audio should be reversed and 3 felt haptic should be reversed. More work is needed to identify which mapping is best for both audio and haptic, or whether additional training time would mitigate the issue. Ultimately, this setting may need to be user-configurable. Future work should also investigate alternative feedback approaches (*e.g.,* the pitch of the speech synthesis could provide directional guidance).

5.2. Feasibility of a Finger-Based Reading Approach

We had expected a finger-based reading approach such as HandSight or Shilkrot *et al.'s* FingerReader [Shilkrot et al. 2014; Shilkrot et al. 2015] to offer many advantages over mobile-based scanners for reading printed text: access to spatial layout information, direct as opposed to sequential access to text on the page, reduced camera framing issues, and, compared to crowdsourced approaches *(e.g.*, [Bigham et al. 2010], BeMyEyes), real-time OCR. However, while we observed some of these advantages, important concerns also arose. Here, we reflect on the feasibility of finger-based reading, incorporating ideas for future work throughout.

5.2.1. Document Layout and Spatial Awareness. A primary motivation for investigating HandSight's finger-based reading approach was to provide users direct access to spatial layout information. Our exploration mode provided audio cues to indicate text, pictures, or white space beneath the user's finger. While 4 participants in Study I commented positively and unprompted on this information, one participant was strongly against the idea, feeling that software that could automatically process a document's layout to extract content would be preferable in most situations. The difficulties encountered by the participant who was removed from our dataset in Study I also highlight an unexpected but important potential for confusion: some users, particularly those who are congenitally blind, may have an inaccurate or incomplete understanding of basic document structures (*e.g.,* columns, margins) simply because they have never encountered them. In that participant's case, he was not familiar with the notion of columns, which led to confusion. Future work should explore the relationship between a user's spatial abilities and their proficiency in exploring a document or responding to finger guidance.

While exploration mode helped participants understand a document's layout (*e.g.,* number of images), distinguishing a gap between paragraphs versus columns was particularly challenging. Both types of gaps were indicated by white space, but participants were frequently unable to determine whether the white space occurred between two paragraphs or between two columns. We intended for paragraph gaps to be distinct from column gaps by the direction in which the finger is moving—vertically for paragraphs or horizontally for columns. However, without sight, many participants tended to move their finger more diagonally, drifting accidentally between paragraphs and columns. This challenge could be addressed by designing cues to identify the horizontal and vertical edges of a block of text.

Finally, we did not evaluate the potential utility of layout information for blind readers. And, arguably, for half the documents we used (the plain text documents), spatial information offered little benefit. The finger-based reading approach may be more beneficial for other types of documents, particularly those with inherent spatial characteristics such as maps or graphs.

5.2.2. Cognitive Load and Physical Effort. Our studies indicate that line-by-line reading incurs high mental and physical effort. The reader must simultaneously attend to directional guidance, document events (*e.g.,* start and end of line), and the synthesized speech content. Study II, in particular, highlighted the increased concentration and physical dexterity required to use HandSight compared to KNFB Reader iOS. This issue of physical effort confirms previous findings from a much smaller study (3 participants) [Shilkrot et al. 2015]. With more practice, HandSight should not require as much effort to use, and, if the technology provides enough

benefit, the need for this practice is not necessarily a barrier to adoption—braille and the Optacon [Millar 2003; Hislop et al. 1983], for example, require extensive practice. However, a multi-session study would be needed to assess just how much practice is needed and how efficiently experienced users can read with a finger-based approach.

5.2.3. Camera Placement. Whether they use crowdsourcing or automated OCR, both mobile document scanning approaches (*e.g.*, KNFB Reader iOS) and body-mounted solutions (*e.g.,* OrCam) require a global image of the document, properly aligned and in focus within the camera's field of view. All participants in Study II reported at least some difficulty with this type of image capture using KNFB Reader iOS, but we had not introduced them to KNFB's document-framing feature. That feature, along with findings from blind photography research (*e.g.* [Vázquez and Steinfeld 2012; Jayant et al. 2011; Cutter and Manduchi 2015]), should reduce the issue. The global image captured by KNFB also allowed for more fluent text-to-speech than with HandSight, which participants valued. At times, however, our own use of KNFB Reader iOS and observations of participants showed that this fluency can provide a false sense of confidence. That is, it is not always clear from the speech output if a part of the document is missing or the application has parsed and played text blocks in the wrong order.

HandSight's finger-mounted camera and direct control over text scanning and speech playback may overcome these issues to some extent, but Study II showed it also introduces new camera placement challenges. For example, participants frequently encountered difficulties tracing lines near the upper and lower margins because the system could not provide directional guidance when no text was visible.

In future work, we plan to explore hybrid methods that may combine a body- or head-mounted camera with a finger-mounted one, potentially overcoming the weaknesses of each and supporting a wider range of reading situations. A body-mounted camera could capture complete documents and allow for efficient, fluent reading using a screen-reader interface and relative exploration of content (*e.g.*, swipes). At the same time, the finger-camera interface could provide knowledge about the document layout, acting as a cursor to quickly search through the content or provide contextual information. It would be useful to compare how well having both types of interaction works compared to only the global, relative interaction or the finger-based interaction.

5.2.4. Physical Design and Social Acceptability. Physical design and social acceptability influence the adoption of wearable technology [Rico and Brewster 2009; Rico and Brewster 2010; Profita et al. 2013]. While our early HandSight prototype is bulky, future versions could be substantially reduced in size since the underlying technology (*i.e.,* the endoscopic camera) is extremely small. Still, whether blind users are interested in wearing a finger-mounted device for accessibility is an open question. Social acceptability could also change how users feel about the haptic line guidance compared to the audio guidance in practice. The majority of users preferred haptic guidance in Study I, but even in future iterations of the physical design, the haptic vibration motors would likely add bulk compared to audio alone. These issues are not unique to HandSight, and the question of where users will feel most comfortable having a camera mounted on their body (if at all) should be explored in future work.

5.2.5. Target Users. While our prototype was designed to support totally blind users, the question of who may benefit most from a finger-based reading approach remains open. Low vision users, for example, may benefit from the direct access and physical gestures that a finger-based reading approach provides, without finding the line tracing as time consuming as for a totally blind reader. We recruited one low vision pilot participant, who experienced no difficulties with describing the layout of the document or with line finding and line tracing. The device could then act as a more portable alternative to closed-circuit television (CCTV) magnifiers, automatically processing the words and providing additional information about the text upon request (*e.g.*, font, spelling). Further investigation, however, is needed to explore this possibility and how it is received compared to commonly used magnifiers.

5.3. Design Iteration

In addition to the future work mentioned above—such as investigating the utility of spatial layout information, conducting a longer-term study, and evaluating HandSight with low-vision users—our findings lead to several design revisions that may improve blind users' experience with HandSight.

We designed the speech interface to adapt to the user's finger speed so that they could easily control the rate of feedback. Some participants liked this feature, but others found it to be choppy when compared to the continuous speech feedback of screen reader software, noting that it was difficult to identify the end of a sentence. More fluid speech feedback and additional audio cues to mark punctuation could ease the reading experience.

An important observation from Study II is that a finger-based reading device should provide an easy way of determining when text is no longer contained within the camera's field of view. Participants occasionally confused situations where the system could not provide guidance (not enough text in the frame) with being correctly centered over the current line and not receiving directional feedback. To address this issue and provide users with more information while reading, document exploration and reading modes could be integrated. In doing so, however, we must take care not to further increase cognitive load and distract from the content of the text.

To reduce the image capture issues seen in Study II, another possibility is to redesign the physical prototype to either move the sensor farther away from the text (as with FingerReader [Shilkrot et al. 2014; Shilkrot et al. 2015], which is on the upper part of the finger) or to use a wide-angle lens. These options could expand the camera's field of view, for example, allowing users to drift farther away from a line before the text is lost.

Finally, our prototypes only allowed users to continue reading forward and did not support backtracking, rereading, or jumping to an arbitrary location in the text. Study I focused on sequential line-tracing guidance, but it would be interesting to implement and evaluate these additional reading actions.

5.4. Limitations

Using an iPad rather than a physical prototype to compare haptic and audio line guidance in Study I was a conscious study design choice, allowing us to bypass technical challenges in implementing a real-time prototype and to focus on the user experience and collect precise line traces. A limitation of this choice, however, is that the experience of reading with a physical prototype and paper may be different. As

well, the font size and document layout for Study I were constrained to two specific formats, which may not fully represent the variety of real-world documents that users may encounter. In Study II, an important limitation of the proof-of-concept prototype is that we assumed that the content of the page was known prior to beginning reading, and constrained the system to allow participants to read text sequentially from left to right and top to bottom. These choices simplified how the system provided finger guidance—it only needed to estimate the finger location on the page and provide upward or downward guidance to return to the last known line. However, these artificial limitations also disregarded some of the potential advantages of a finger-based reading approach, such as re-reading or jumping to arbitrary locations. We also asked participants to hold their hand in a specific position for Study II, constraining their natural behavior when using a device such as ours. Study II was not meant to offer a controlled comparison of KNFB Reader iOS and HandSight, but limitations even for gathering exploratory feedback include that we did not evaluate the document framing feedback of KNFB Reader iOS, and that only one participant had previous experience with KNFB Reader iOS (all 4 had used finger-based reading). While a more controlled comparison is thus needed, it is important to note that participants still identified many strengths of KNFB Reader iOS. Finally, while we focused on blind readers, it would be interesting to expand the evaluation of finger-based reading to users with a wider range of visual abilities.

6. CONCLUSION

We conducted an in-depth study with 19 blind participants comparing audio and haptic cues for directional guidance to support finger-based reading. Our findings showed similar performance and user preference between the two types of guidance, although audio resulted in significantly more accurate line tracing for some tasks. Subjective feedback was split, but suggests that haptic guidance may be slightly preferred. In addition, our findings highlighted general strengths and weaknesses of a finger-based reading approach, such as improved understanding of a document's layout and the difficulty encountered by blind users in accurately tracing a line of text with a finger. In follow-up sessions where 4 of the participants used a proof-of-concept finger-reading prototype as well as KNFB Reader iOS, the mobile scanner was seen as offering a more fluent reading experience. Ultimately, a finger-based reading approach may be best suited to material that is inherently spatial, such as maps or graphs, whereas existing applications that capture a global image of the document for text-to-speech (*e.g.*, KNFB Reader iOS) may be preferred for text-heavy material. Future work should investigate this possibility, as well as assess the potential of finger-based reading for low-vision users, for whom precise directional finger guidance may not be necessary.

ACKNOWLEDGMENTS

We thank our participants and the Maryland State Library for the Blind and Physically Handicapped. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

REFERENCES

Chieko Asakawa, Hironobu Takagi, Shuichi Ino, and Tohru Ifukube. 2003. Maximum listening speeds for the blind. In *Proceedings of the 2003 International Conference on Auditory Display*. 276–279.

Jeffrey P. Bigham et al. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the*

23nd annual ACM symposium on User interface software and technology - UIST '10. New York, New York, USA: ACM Press, 333–342. DOI:http://dx.doi.org/10.1145/1866029.1866080

Jeffrey P. Bigham, Richard E. Ladner, and Yevgen Borodin. 2011. The design of human-powered access technology. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '11*. New York, New York, USA: ACM Press, 3. DOI:http://dx.doi.org/10.1145/2049536.2049540

Nicoletta Di Blas, Paulo Paolini, and Marco Speroni. 2004. Useable accessibility to the web for blind users. In *Proceedings of the 8th ERCIM Workshop: User Interfaces for All*.

J.C. Bliss. 1969. A relatively high-resolution reading aid for the blind. *Man-Machine Syst. IEEE Trans.* 10, 1 (March 1969), 1–9. DOI:http://dx.doi.org/10.1109/TMMS.1969.299874

Yevgen Borodin, Jeffrey P. Bigham, Glenn Dausch, and I. V. Ramakrishnan. 2010. More than meets the eye: a survey of screen-reader browsing strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) - W4A '10*. New York, New York, USA: ACM Press, 1. DOI:http://dx.doi.org/10.1145/1805986.1806005

Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013. Visual challenges in the everyday lives of blind people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. New York, New York, USA: ACM Press, 2117–2126.

Karl Burklen. 1932. *Touch reading of the blind*, New York: American Foundation for the Blind.

M. Capp and P. Picton. 2000. The optophone: an electronic blind aid. *Eng. Sci. Educ. J.* 9, 3 (2000), 137–143. DOI:http://dx.doi.org/10.1049/esej:20000306

Xiangrong Chen and A.L. Yuille. 2004. Detecting and reading text in natural scenes. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. II–366–II–373 Vol.2. DOI:http://dx.doi.org/10.1109/CVPR.2004.1315187

Franklin S. Cooper, Jane H. Gaitenby, and Patrick W. Nye. 1983. *Evolution of reading machines for the blind: Haskins Laboratories' research as a case history*, Haskins Laboratories.

Andrew Crossan and Stephen Brewster. 2008. Multimodal trajectory playback for teaching shape information and trajectories to visually impaired computer users. *ACM Trans. Access. Comput.* 1, 2 (October 2008), 12:1–12:34. DOI:http://dx.doi.org/10.1145/1408760.1408766

Michael Cutter and Roberto Manduchi. 2015. Towards Mobile OCR: How To Take a Good Picture of a Document Without Sight. *Proc. ACM Symp. Doc. Eng. ACM Symp. Doc. Eng.* 2015 (January 2015), 75–84.

EE Fournier D'Albe. 1914. On a type-reading optophone. *Proc. R. Soc. London. Ser. A* 90, 619 (1914), 373–375.

Sarah L. Dowhower. 1989. Repeated reading: research into practice. *Read. Teach.* 42, 7 (1989), pp. 502–507.

Leah Findlater et al. 2015. Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 383–384.

Vincent Gaudissart, Silvio Ferreira, Céline Thillou, and Bernard Gosselin. 2004. SYPOLE: mobile reading assistant for blind people. In *9th Conference Speech and Computer (SPECOM)*. St. Petersburg, Russia.

Nicholas A. Giudice, Hari Prasath Palani, Eric Brenner, and Kevin M. Kramer. 2012. Learning non-visual graphical information using a touch-based vibro-audio interface. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*. New York, New York, USA: ACM Press, 103.

Louis H. Goldish and Harry E. Taylor. 1973. The Optacon: a valuable device for blind persons. *New Outlook Blind* (November 1973).

Susumu Harada, Hironobu Takagi, and Chieko Asakawa. 2011. On the Audio Representation of Radial Direction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. New York, NY, USA: ACM, 2779–2788. DOI:http://dx.doi.org/10.1145/1978942.1979354

Richard Hartley and Andrew Zisserman. 2003. *Multiple View Geometry in Computer Vision*, Cambridge University Press.

Sharon A. Haymes, Alan W. Johnston, and Anthony D. Heyes. 2002. Relationship between vision impairment and ability to perform activities of daily living. *Ophthalmic Physiol. Opt.* 22, 2 (March 2002), 79–91. DOI:http://dx.doi.org/10.1046/j.1475-1313.2002.00016.x

David W. Hislop, B.L. Zuber, and John L. Trimble. 1983. Characteristics of reading rate and manual scanning patterns of blind Optacon readers. *Hum. Factors J. Hum. Factors Ergon. Soc.* 25, 4 (August 1983), 379–389.

David W. Hislop, B.L. Zuber, and John L. Trimble. 1985. Text-scannning patterns of blind readers using Optacon and braille. *J. Rehabil. Res. Dev.* (1985), 54–65.

Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* (1979), 65–70.

Bill Holton. 2014. KNFB Reader for iOS: does this app live up to all the hype? *AFB AccessWorld Mag.* 15, 11 (2014).

Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. ASSETS '11. New York, NY, USA: ACM, 203–210. DOI:http://dx.doi.org/10.1145/2049536.2049573

Shaun K. Kane, Brian Frey, and Jacob O. Wobbrock. 2013. 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. New York, NY, USA: ACM, 347–350. DOI:http://dx.doi.org/10.1145/2470654.2470704

R. Keefer, Yan Liu, and N. Bourbakis. 2013. The development and evaluation of an eyes-free interaction model for mobile reading devices. *Human-Machine Syst. IEEE Trans.* 43, 1 (2013), 76–91. DOI:http://dx.doi.org/10.1109/TSMCA.2012.2210413

Deborah Kendrick. 2008. A picture becomes a thousand words: an evaluation of the knfbReader mobile. *Am. Found. Blind AccessWorld Mag.* 9, 6 (2008).

Georgios Kouroupetroglou and Dimitrios Tsonos. 2008. Multimodal accessibility of documents. In *Advances in Human-Computer Interaction*. Vienna, Austria: I-Tech Education and Publishing, 451–470.

Jonathan Lazar, Aaron Allen, Jason Kleinman, and Chris Malarkey. 2007. What frustrates screen reader users on the web: a study of 100 blind users. *Int. J. Hum. Comput. Interact.* 22, 3 (2007), 247–269. DOI:http://dx.doi.org/10.1080/10447310709336964

Betty Ann Levy. 2001. Text processing: Memory representations mediate fluent reading. *Perspect. Hum. Mem. Cogn. aging Essays honour Fergus Craik* (2001), 83–98.

Roberto Manduchi and James M. Coughlan. 2014. The last meter: blind visual guidance to a target. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'14)*. To appear.

Susanna Millar. 2003. *Reading by touch*, Routledge.

S. Mori, C.Y. Suen, and K. Yamamoto. 1992. Historical review of OCR research and development. *Proc. IEEE* 80, 7 (July 1992), 1029–1058. DOI:http://dx.doi.org/10.1109/5.156468

Elizabeth D. Mynatt and Gerhard Weber. 1994. Nonvisual presentation of graphical user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94*. New York, New York, USA: ACM Press, 166–172.

National Center for Health Statistics. 2012. *National Health Interview Survey 2012 Data Release*,

National Federation of the Blind Jernigan Institute. 2009. The braille literacy crisis in america: facing the truth, reversing the trend, empowering the blind. (2009).

Nicolas Noble and Benoıt Martin. 2006. Shape discovering using tactile guidance. In *Proceeding of the 6th International Conference EuroHaptics*.

Uran Oh, Shaun K. Kane, and Leah Findlater. 2013. Follow that sound: using sonification and corrective verbal feedback to teach touchscreen gestures. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. Article 13.

Peter Parente and Gary Bishop. 2003. BATS: The Blind Audio Tactile Mapping System. *Proc. ACM South East. Conf.* (2003).

Beryl Plimmer, Peter Reid, Rachel Blagojevic, Andrew Crossan, and Stephen Brewster. 2011. Signing on the tactile line: a multimodal system for teaching handwriting to blind children. *ACM Trans. Comput. Interact.* 18, 3 (August 2011), 17:1–17:29. DOI:http://dx.doi.org/10.1145/1993060.1993067

Benjamin Poppinga, Charlotte Magnusson, Martin Pielot, and Kirsten Rassmus-Gröhn. 2011. TouchOver map. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*. New York, New York, USA: ACM Press, 545.

The Princeton Review. 2000. *ITBS preparation and practice workbook, Glencoe language arts, grade 8*,

New York, NY: McGraw-Hill/Glencoe.

Halley P. Profita et al. 2013. Don't mind me touching my wrist. In *Proceedings of the 17th annual international symposium on International symposium on wearable computers - ISWC '13*. New York, New York, USA: ACM Press, 89.

Julie Rico and Stephen Brewster. 2009. Gestures all around us: user differences in social acceptability perceptions of gesture based interfaces. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '09*. New York, New York, USA: ACM Press, 64. DOI:http://dx.doi.org/10.1145/1613858.1613936

Julie Rico and Stephen Brewster. 2010. Usable gestures for mobile interfaces. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. New York, New York, USA: ACM Press, 887. DOI:http://dx.doi.org/10.1145/1753326.1753458

Ruby Ryles. 1996. The impact of braille reading skills on employment, income, education, and reading habits. *J. Vis. Impair. Blind.* 90 (1996), 219–226.

R. Safabakhsh and S. Khadivi. 2000. Document skew detection using minimum-area bounding rectangle. In *Proceedings International Conference on Information Technology: Coding and Computing (Cat. No.PR00540)*. IEEE Comput. Soc, 253–258. DOI:http://dx.doi.org/10.1109/ITCC.2000.844226

Huiying Shen and JamesM. Coughlan. 2012. Towards a real-time system for finding and reading signs for visually impaired users. In Klaus Miesenberger, Arthur Karshmer, Petr Penaz, & Wolfgang Zagler, eds. *Computers Helping People with Special Needs SE - 7*. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 41–47. DOI:http://dx.doi.org/10.1007/978-3-642-31534-3_7

Roy Shilkrot, Jochen Huber, Connie Liu, Pattie Maes, and Nanayakkara Suranga Chandima. 2014. FingerReader: a wearable device to support text reading on the go. *CHI '14 Ext. Abstr. Hum. Factors Comput. Syst.* , Vi (2014), 2359–2364. DOI:http://dx.doi.org/10.1145/2559206.2581220

Roy Shilkrot, Jochen Huber, Meng Ee Wong, Pattie Maes, and Suranga Nanayakkara. 2015. FingerReader: a wearable device to explore printed text on the go. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '15)*. 2363–2372.

Lee Stearns et al. 2014. 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 (ACVR'14) in Conjunction with the European Conference on Computer Vision (ECCV'14)*. Zurich, Switzerland: Springer, 615–631.

Jing Su, Alyssa Rosenzweig, Ashvin Goel, Eyal de Lara, and Khai N. Truong. 2010. Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services - MobileHCI '10*. 17–26. DOI:http://dx.doi.org/10.1145/1851600.1851606

J.L. Toennies, J. Burgner, T.J. Withrow, and R.J. Webster. 2011. Toward haptic/aural touchscreen display of graphical mathematics for the education of blind students. In *2011 IEEE World Haptics Conference*. IEEE, 373–378.

Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '12*. New York, New York, USA: ACM Press, 95. DOI:http://dx.doi.org/10.1145/2384916.2384934

Kai Wang, B. Babenko, and S. Belongie. 2011. End-to-end scene text recognition. In *Computer Vision (ICCV), 2011 IEEE International Conference on*. 1457–1464. DOI:http://dx.doi.org/10.1109/ICCV.2011.6126402

Xing-Dong Yang, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2012. Magic finger: always-available input through finger instrumentation. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. UIST '12. New York, NY, USA: ACM, 147–156. DOI:http://dx.doi.org/10.1145/2380116.2380137

Koji Yatani, Nikola Banovic, and Khai Truong. 2012. SpaceSense: representing geographical information to visually impaired people using spatial tactile feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. New York, NY, USA: ACM, 415–424. DOI:http://dx.doi.org/10.1145/2207676.2207734

Koji Yatani and Khai Nhut Truong. 2009. SemFeel: a user interface with semantic tactile feedback for mobile touch-screen devices. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*. UIST '09. New York, NY, USA: ACM, 111–120. DOI:http://dx.doi.org/10.1145/1622176.1622198

APPENDIX A

In this appendix we list the text of the documents used in Studies I and II, along with associated comprehension questions. We adapted six test documents from a Grade 8 Iowa Test of Basic Skills practice book [Princeton Review 2000]. The original text was modified slightly for length and to ensure clarity with our speech synthesis engine (*e.g.,* removing unnecessary proper nouns). We created three additional training documents of a similar length and reading level, as well as a two-column magazine document for testing KNFB Reader iOS, using documents adapted from articles in USA Today.

Study I, Training Document (plain, both conditions):

Scientists counting Antarctica's emperor penguins from space have found twice as many of them as expected. Using high-resolution satellite images to study each of 44 colonies around the coastline of Antarctica, experts said Friday that they put the total emperor penguin population at 595 thousand, or roughly double previous estimates of 270 thousand to 350 thousand. Seven of the colonies had never been seen before.

Satellite technology was a boon for researchers; visiting dozens of remote colonies in temperatures as low as minus 58 degrees is expensive, dangerous and time-consuming. With their distinctive black and white plumage, emperor penguins stand out against the snow, making them clearly visible on satellite images.

Study I, Test Document 1 (plain, first condition):

People have used coins as a means of exchange for thousands of years. Valued for their craftsmanship and purchasing power, coins have been collected in great numbers throughout history and buried for safekeeping. Because stores of coins gathered and hidden in this manner lie untouched for many years, they can reveal a great deal about a given culture.

Coins are useful in revealing many aspects of a culture. They can provide clues about when a given civilization was wealthy and when it was experiencing a depression. Wealthy nations tend to produce a greater number of coins made from richer materials. The distribution of coins can also reflect the boundaries of an empire and the trade relationships within it. Roman imperial gold coins found in India, for example, indicate the Romans purchased goods from the East.

The way the coins themselves are decorated sometimes provides key information about a culture. Many coins are stamped with a wealth of useful historical evidence, including portraits of political leaders, important buildings and sculptures, mythological and religious figures, and useful dates. Some coins, such as many from ancient Greece, can be considered works of art themselves and reflect the artistic achievement of the civilization as a whole.

Information gathered from old coins by historians is most useful when placed alongside other historical documents, such as written accounts or data from archeological digs. Combined with these other pieces of information, coins can help historians reconstruct the details of lost civilizations.

Comprehension Questions:

1. Which of the following do coins reveal about a civilization?
    a. The average cost of clothing
    b. Information about its economy
    c. Its farming techniques
2. What is the main idea of the passage?
    a. How difficult it is to find old coins
    b. How coins reflect the artistic achievements of a culture
    c. How coins can tell us about ancient civilizations

Study I, Test Document 2 (magazine, first condition):

Despite the stubborn, widespread opinion that animals don't feel emotions in the same way that humans do, many animals have been observed to demonstrate a capacity for joy. People have often seen animals evincing behavior that can only be taken to mean they are pleased with what life has brought them in that particular moment.

A chimpanzee named Nim was raised by a human family for the first year and a half of his life. After that time, Nim was separated from them for two and a half years. On the day that Nim was reunited with his human family, he smiled, shrieked, pounded the ground, and looked from one member of the family to the next. Still smiling and shrieking, Nim went around hugging each member of the family. He played with and groomed each member of the family for almost an hour before the family had to leave. People who were familiar with Nim's behavior said they had never seen him smile for such a long period of time.

Comprehension Questions:

1. What is the purpose of the second paragraph?
    a. To criticize Nim's human family for abandoning him
    b. To show how well Nim's human family treated him
    c. To demonstrate that animals have the ability to feel joy
2. Why did Nim shriek and pound the ground?
    a. He was overjoyed to see the family again.
    b. He was hungry and wanted to be fed.
    c. He was angry with the family for leaving him.

Study I, Test Document 3 (plain, second condition):

Born in Spanish Harlem in the late 1950s, Raphael Sanchez learned at an early age to listen to the many voices of the city. It was as a boy in Harlem that he developed the powers of observation that would later make his writing truly great. In the 1970s, Raphael went to Columbia University, where he was exposed to a literary tradition. While his university education gave his writing new depth, the raw energy of the streets has always served as the primary fuel for his writing. This is what gives his works passion and power.

Raphael once told me that in order to escape from life he turns to books, and in order to escape from books he turns to life. It is this balance of the sights, sounds, and smells of the street with the perspective gained from his formal education that has made Raphael popular with both critics and regular readers alike.

For those of us who have read and admired his work, it seems natural that Raphael has won so many awards. He deserves them, and his humility in accepting them has been refreshing. When he received the Writer's Quill Award two weeks ago, for example, he told the audience, "This award is not really mine. It belongs to all the million things that have inspired me.

"That is the kind of man I am introducing to you this evening. He is a man who has been inspired by a million things, and he is a man who has provided inspiration to a million people. Ladies and gentlemen, it is my great honor to present to you, Raphael Sanchez.

Comprehension Questions:

1. Which of these best describes why Raphael Sanchez's writing is so popular with critics and regular readers?
    a. It has won the Writer's Quill Award.
    b. It reflects both scholarship and city experience.
    c. It is based on his experiences at Columbia University.
2. What does Raphael Sanchez mean when he says, "This award is not really mine"?
    a. He owes everything to the people and things that inspired him.
    b. He does not believe in the value of awards.
    c. He feels Columbia University should be given an award too.

Study I, Test Document 4 (magazine, second condition):

In the 1800s, most geologists thought the sea floor was a lifeless expanse of mud, sediment, and the decaying remains of dead organisms. They thought that, with the exception of some volcanic islands, the bottom of the sea had no major geographic features, such as peaks or valleys.

In the mid-nineteenth century, ships depth-sounding the ocean floor with sonar for a transatlantic telegraph cable made some interesting discoveries. To geologists' surprise, the ocean floor was found to be made up of long mountain ranges and deep valleys and troughs. Another surprise finding in the Atlantic was the existence of basalt, a volcanic rock thought only to exist in the Pacific Ocean. The presence of basalt in the Atlantic was a clue that volcanic activity occurs at the bottom of the sea. This and other discoveries, many of them accidental in the beginning, were signals to geologists that their knowledge of the sea floor was very limited.

Comprehension Questions:

1. The discovery of basalt in the Atlantic Ocean suggested that
    a. Iron, zinc, and gold would be found nearby.
    b. Geologists still had much to learn about the ocean floor.
    c. The Atlantic was deeper than previously believed.
2. How did ships in the mid-nineteenth century measure the ocean's depth?
    a. By sending down scuba divers
    b. By bouncing sound waves off the sea bottom
    c. By photographing the sea floor with special cameras

Study II, Training Document 1 (plain, HandSight):

When Mary Smallenburg opened a package from her mother to find cereal and ramen noodles, she burst into tears. Without it, she wouldn't be able to feed her four children. It got to the point where I opened my pantry and there was nothing. Nothing. What was I going to feed my kids? Smallenburg says, adjusting a bag of fresh groceries on her arm.

Smallenburg's family is one of 50 military families that regularly visit the Lorton Community Action Center food bank. Volunteers wave a familiar hello as she walks in the door. None of what we have been through has been expected, Smallenburg says. Three of her four children have special needs and her husband is deployed in Korea. The last few months actually, coming here has been a godsend.

Nationwide, 25 percent of military families, 620,000 households, need help putting food on the table, according to a study by Feeding America, a network of 200 food banks. The results are alarming, says Bob Aiken, chief executive officer of Feeding America. It means that people in America have to make trade-offs. They have to pick between buying food for their children or paying for utilities, rent and medicine.

One in seven Americans, 46 million people, rely on food pantries and meal service programs to feed themselves and their families, the study found.

Study II, Test Document 1 (plain, HandSight):

Henry Ford and his Model T automobile changed the face of America. His horseless carriage contributed to a movement from rural to urban and to the development of an industrial economy.

In 1903, Ford Motor Company was officially formed, and in 1908, Ford announced the birth of the Model T. He told the nation, I will build a car for the great multitude. This was a bold announcement, since most manufacturers planned to build only luxury cars for the very wealthy.

His idea worked. Ford's Model T was a hit with the American public, and demand grew with each passing year. In the course of nineteen years, around fifteen million Model T cars were sold in the United States, nearly one million in Canada, and another 250,000 in Great Britain. All told, these numbers equaled half the total number of automobiles manufactured in the world at that time.

The methods of production Ford used were revolutionary. Ford's assembly line could churn out the frame of a Model T in about six hours. This high rate of speed was made possible by the division of labor. Instead of one person controlling production from start to finish, the labor was divided into smaller and smaller tasks, with each person performing the same task all day long.

By 1927 the era of the Model T was coming to a close. General Motors, a major competitor, was producing better cars for only slightly more money. Customers with an eye for new styles just didn't see the appeal of the plain Model T.

Comprehension Questions:

1. According to the passage, why was the Model T more popular than other cars available at the same time?
    a. It looked like a buggy.
    b. It was more spacious.
    c. It was less expensive.
    d. It was more stylish.
2. Which of the following best describes Ford according to the article?
    a. A poor businessman
    b. A visionary
    c. A follower
    d. A great metal worker
3. What led to the downfall of the Model T?
    a. It was not very well made.
    b. Its price went up.
    c. Other competition emerged.
    d. Many of Ford's workers quit their jobs.

Study II, Training Document 2 (plain, KNFB Reader iOS):

Here's a tip. Don't stress over tipping.

Restaurant tips are more modest in Europe than in America. In most places, 10 percent is a big tip. If your bucks talk at home, muzzle them on your travels. As a matter of principle, if not economy, the local price should prevail. Please believe me, tipping 15 percent or 20 percent in Europe is unnecessary, if not culturally ignorant.

Virtually anywhere in Europe, you can do as the Europeans do and, if you're pleased with the service, add a euro or two for each person in your party. In very touristy areas, some servers have noticed the American obsession with overtipping, and might hope for a Yankee-size tip. But the good news is that European servers and diners are far more laid-back about all this than we are. The stakes are low, and it's no big deal if you choose the wrong amount. And note that tipping is an issue only at restaurants that have waiters and waitresses. If you order your food at a counter, don't tip.

At table-service restaurants, the tipping etiquette and procedure vary slightly from country to country. But in general, European servers are well paid, and tips are considered a small bonus, to reward great service or for simplicity in rounding the total bill to a convenient number. In many countries, 5 percent to 10 percent is sufficient.

Study II, Test Document 2 (plain, KNFB Reader iOS):

A clone is a life form engineered in a lab environment to be identical to another, through a process of asexual, or nonsexual, reproduction. This process of creating a new life form, called genetic engineering, can be useful in creating individuals of a given species that represent the best possible genetic traits of that species. People who work with plants have long used cloning techniques to create better strains of trees, fruits, and vegetables. The Macintosh apple, for example, was created by cloning techniques, and it supposedly represents the best qualities of all apple types.

In July of 1996, a group of Scottish scientists made a breakthrough by successfully cloning a sheep from the cells of another adult sheep. After scraping cells from the udder of one sheep, the scientists introduced the nucleus of one of these cells into the unfertilized eggs of a different sheep. Then, they placed the egg, which had begun to divide, into the uterus of a third sheep. The result was Dolly, a healthy sheep who was born in the natural way from the third sheep. Dolly was almost identical to the sheep from whose skin cells she had been formed.

In 1997 Dolly's story was widely publicized in the media, and her existence resparked a continuing debate about the use of cloning techniques on humans. Some people claim that genetic engineering should not be used on humans under any circumstances. Others urge slowness. They insist that if genetic engineering is to be used, there are many questions that need to be answered first.

Comprehension Questions:

1. What event revived the debate about cloning?
    a. The cloning of plants
    b. The cloning of a sheep
    c. The cloning of amphibians
    d. The future plan to clone human organs
2. According to the passage, how many sheep did it take to produce Dolly?
    a. One
    b. Two
    c. Three
    d. Four
3. In paragraph 1, the author mentions the Macintosh apple as an example of
    a. a case when cloning produced an improved organism.
    b. a case when cloning failed.
    c. a case when cloning created a controversy.
    d. a case when cloning went too far.

Study II, Test Document 3 (magazine, KNFB Reader iOS):

Let them sleep! That's the message from the nation's largest pediatrician group, which, in a new policy statement, says delaying the start of high school and middle school classes to 8:30 a.m. or later is an effective countermeasure to chronic sleep loss and the epidemic of delayed, insufficient, and erratic sleep patterns among the nation's teens.

Multiple factors, including biological changes in sleep associated with puberty, lifestyle choices, and academic demands, negatively impact teens' ability to get enough sleep, and pushing back school start times is key to helping them achieve optimal levels of sleep, 8 and a half to 9 and a half hours a night, says the American Academy of Pediatrics statement, released Monday and published online in Pediatrics.

Just 1 in 5 adolescents get nine hours of sleep on school nights, and 45 percent sleep less than eight hours, according to a 2006 poll by the National Sleep Foundation (NSF).

As adolescents go up in grade, they're less likely with each passing year to get anything resembling sufficient sleep, says Judith Owens, director of sleep medicine at Children's National Medical Center in Washington, D.C., and lead author of the AAP statement. By the time they're high school seniors, the NSF data showed they were getting less than seven hours of sleep on average.

Chronic sleep loss in children and adolescents can, without hyperbole, really be called a public health crisis, Owens says.

APPENDIX B

In this appendix we list the text of the subjective questionnaires administered in Studies I and II. For ease of use questions, the choices were (1) Very difficult, (2) Somewhat difficult, (3) Neutral, (4) Somewhat easy, or (5) Very easy. For comparison questions, the choices were (1) Much worse, (2) Somewhat worse, (3) About the same, (4) Somewhat better, or (5) Much better.

Study I, after each directional guidance condition:

1. *How easy or difficult was it to follow a line of text with your finger? Why?*
2. *How easy or difficult was it to read and understand the text given this feedback? Why?*

3. *Do you feel like the feedback direction was correct, or did if feel backwards to you?*
4. *Do you have any other comments about what was good or bad about this type of feedback?*

Study I, end of study:

1. *Overall, how easy or difficult was it to find the beginning of the text?*
2. *How easy or difficult was it to find the beginning of each line?*
3. *How easy or difficult was it to notice the end of a line?*
4. *How easy or difficult was it to notice the end of a paragraph?*
5. *How easy or difficult was it to find the beginning of the next column?*
6. *You have tried two different types of feedback. Which did you prefer more? Why?*
7. *Overall, how was your experience of our system compared to how you would normally read braille? Why?*
8. *Overall, how was your experience of our system compared to how you would normally use a screen reader? Why?*
9. *Overall, how was your experience of our system compared to how you would normally read printed documents? Why?*
10. *Do you have any questions, suggestions for improvement, or other comments?*

Study II, after HandSight tasks:

1. *How easy or difficult was it to find the beginning of the text?*
2. *How easy or difficult was it to find the beginning of each line?*
3. *How easy or difficult was it to notice the end of a line?*
4. *How easy or difficult was it to notice the end of a paragraph?*
5. *How easy or difficult was it to follow a line of text with your finger?*
6. *How easy or difficult was it to understand the feedback?*
7. *Overall, how easy or difficult was it to read and understand the text given this feedback?*
8. *Overall, how was your experience with the app compared to how you would normally read braille?*
9. *Overall, how was your experience with the app compared to how you would normally read printed documents?*
10. *Do you have any other comments about what was good or bad about this type of feedback?*

Study II, after KNFB Reader iOS tasks:

1. *Overall, how easy or difficult was it to read and understand the text using the app?*
2. *What, if anything, did you like about using the app?*
3. *What, if anything, did you dislike about using the app?*
4. *Overall, how was your experience with the app compared to how you would normally read braille documents?*
5. *Overall, how was your experience with the app compared to how you would normally read printed documents?*
6. *Overall, how was your experience with the app compared to reading with HandSight?*
7. *Do you have any questions, suggestions for improvement, or other comments about the KNFB Reader app?*

Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera

Lee Stearns[1], Victor DeSouza[1], Jessica Yin[3], Leah Findlater[2,5], Jon E. Froehlich[1,4]

[1]Department of Computer Science
[2]College of Information Studies
University of Maryland, College Park

[3]Poolesville High School
Poolesville, MD

[4]Computer Science and Engineering
[5]Human Centered Design and Engineering
University of Washington

lstearns@umd.edu, vdesouza@umd.edu, jessica.y.phs@gmail.com, leahkf@umd.edu, jonf@cs.umd.edu

ABSTRACT

Recent technical advances have enabled new wearable augmented reality (AR) solutions that can aid people with visual impairments (VI) in their everyday lives. Here, we investigate an AR-based magnification solution that combines a small finger-worn camera with a transparent augmented reality display (the Microsoft Hololens). The image from the camera is processed and projected on the Hololens to magnify visible content below the user's finger such as text and images. Our approach offers: (i) a close-up camera view (similar to a CCTV system) with the portability and processing power of a smartphone magnifier app, (ii) access to content through direct touch, and (iii) flexible placement of the magnified image within the wearer's field of view. We present three proof-of-concept interfaces and plans for a user evaluation.

CCS Concepts

• Human-centered computing → Accessibility technologies • Computing methodologies → Mixed / augmented reality

Keywords

Visually impaired; wearable computing; assistive technology; augmented reality; reading printed text; magnification

1. INTRODUCTION

Augmented reality (AR) has the potential to enhance the accessibility of the physical world for people with visual impairments (VI). One important emerging application area is the magnification and enhancement of visual information. Traditional methods such as closed-circuit television (CCTV) systems and other desktop video magnifiers [2] allow users with some residual vision to magnify and display content on a large screen. Many users also rely on magnifying glasses or smartphone-based magnifiers [1], which are more portable but offer a limited visual area and require the user to hold a device while reading. In contrast, a wearable AR system is portable and always available, provides a perceptually large display, and can allow for magnified output that is visually co-located with or even overlaid on top of the original content within the wearer's field of view.

In this poster paper, we introduce a novel AR magnification approach that combines a finger-worn camera to capture visual content with a transparent head-worn display for displaying a magnified and/or adjusted view of that content—in our case, using a *Microsoft Hololens* device [5]. Two previous approaches to magnifying information on a head-worn display have used video see-through displays where a video feed from a head-worn camera is displayed on screens in front of the user's eyes along with computer-generated content [3, 8]. In contrast, our approach does not restrict the user's field of view to only the camera view, allows for small finger movements to select the content to magnify (as opposed to head movements), and allows for 3D virtual content to be overlaid within the physical world. This lattermost characteristic is enabled by the Hololens API, which provides robust 3D mapping and motion tracking to make AR content appear as if it is fixed in place in the physical world.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ASSETS '17, October 29-November 1, 2017, Baltimore, MD, USA
© 2017 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-4926-0/17/10.
https://doi.org/10.1145/3132525.3134812

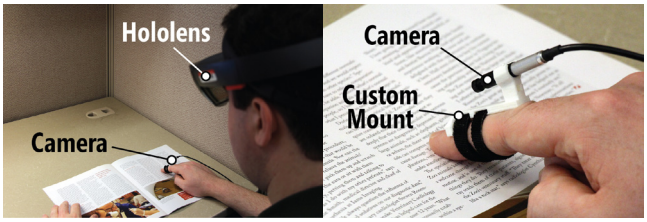


Figure 1. Prototype Hololens and finger-camera system: (a) reading a magazine article; (b) close-up view of the camera.

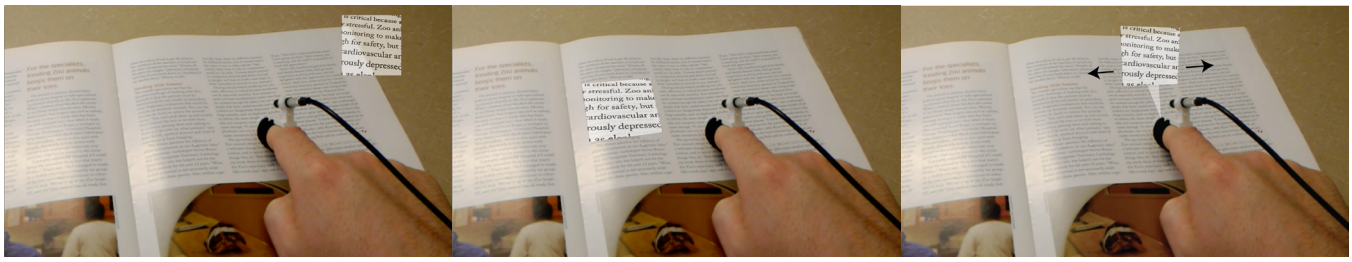
To date, we have implemented a proof-of-concept system with three interface variants for sensing, processing, and displaying magnified and/or adjusted images (*e.g.*, pictures, text). We describe these interfaces and associated design considerations, discuss plans for future user evaluations, and enumerate open questions for assistive AR applications. Our contributions include: (i) a preliminary investigation of a finger-mounted camera with an AR display for viewing printed or other non-tactile information and (ii) a proof-of-concept AR system with three distinct interface options for displaying magnified visual content.

2. PROTOTYPE HARDWARE

Our prototype system (Figure 1) includes a head-worn Microsoft Hololens unit and an Awaiba NanEye Idule camera worn on the index finger via a custom ring. Similar cameras have been applied for automated reading of printed text [6, 7] but here we use it solely as a digital magnifier. Currently, the camera connects to a laptop, which processes the video feed and communicates with the Hololens over Wi-Fi. While the Hololens is bulky and relatively heavy, its advanced capabilities and high-resolution display allowed us to quickly prototype and explore novel AR interfaces. In the future, we envision a fully wearable, lightweight system.

3. INTERFACE DESIGNS

To explore the magnification design space afforded by the combination of a head-worn display and finger-worn camera, we designed three interface options that vary in how the magnified AR view is displayed (Figure 2). Similar to *ForeSee* [8], we also implemented and explored real-time image processing techniques



(a) Fixed position on 2D AR display (*i.e.*, fixed within the wearer's field of view).

(b) Fixed position on 3D surface or midair, in this case rendered flat on the magazine page.

(c) Dynamic finger tracking, where magnified view follows and hovers above the finger.

Figure 2. The three viewport positions in our prototype system, which vary in how they present the magnified view to the user.

such as color and contrast changes to enhance legibility for low-vision users.

Design 1: Fixed Position on 2D AR Display. Our first design places the AR content at a fixed position within the user's field of vision, no matter how they move their head. We envision this design working similarly to other wearable heads-up displays such as Google Glass [4] but allowing for greater customization of the size and position of the magnified information. These parameters can be adjusted to account for user preference or vision level (*e.g.*, at the periphery for users with macular degeneration). This design is simple to implement and easy to understand but also partially obstructs a portion of the user's vision at all times.

Design 2: Fixed Position on 3D Surface. A second design places the display at a fixed position in 3D space, creating a virtual screen (the magnified image) on any surface the user chooses (*e.g.*, on a table or wall or on the document that the user is reading) or floating in midair. This design has the advantage of not occupying a part of the user's vision, instead staying at a fixed location in the world no matter how the user moves. This type of display may also more seamlessly integrate the virtual and physical worlds with content overlaid alongside or even in place of the original text in the physical world. However, this design requires more complex processing and user interactions, and since the virtual display is partially transparent, it may not work well for all locations.

Design 3: Dynamic Finger Tracking. In our third design, the display is positioned floating above the finger as it glides along the page. This design helps prevent disorientation by keeping the display as close as possible to the touched location so that the user can more easily view both. The interaction is similar to holding a physical magnifying glass or smartphone magnifier except that the user can do so at a greater distance and does not need to hold a physical device in hand, leaving the fingers free to touch the page. This design is the most complex, requiring reliable finger tracking and continual updating as the user moves, and it is possible that the motion of the display may distract from the content.

4. DISCUSSION AND OPEN QUESTIONS

Below we discuss open questions related to our system's design, plans for user evaluations, and our long-term vision for AR-based assistance for low-vision users.

The Hololens allowed us to easily prototype and experiment with novel AR designs; however, it has limitations that should be addressed for future versions of the system—specifically its size and relatively narrow display area centered in the user's field of view. Similarly, our finger-worn camera hardware is somewhat large fore regular use, and more importantly is not fully wearable—it currently relies on a separate computer for power and streaming images to the display.

While we have focused primarily on how and where to display the magnified content, designing appropriate user interactions is also important. For each of our interface designs, the user should be able to customize the magnification level, position, text processing, and other settings. Interacting with the system could be done via hand gestures or speech recognition (both are common with the Hololens). Future usability testing here is necessary to determine an appropriate interaction model.

To assess our three AR designs, we plan to conduct an evaluation with low-vision participants comparing the designs in terms of usability and comprehension. We will also compare finger-worn and head-worn cameras to explore the tradeoffs of each and compare our AR-based magnification approach against the current status quo—a handheld smartphone magnifier.

We also envision a future AR system that goes beyond simple magnification. For example, if combined with optical character recognition, the system could change fonts (*e.g.*, [8]) or read the text aloud as needed using text-to-speech. Content could be dynamically enhanced and placed over the original, although for legibility, the AR overlays would need to be opaque rather than the semi-transparent holograms that are currently possible with the Hololens. Future systems could also include touchscreen text and image manipulation features such as highlighting or copy and paste, potentially broadening the appeal to users with any level of vision.

5. ACKNOWLEDGMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

6. REFERENCES

[1] Apple. iOS Accessibility Features: *https://www.apple.com/accessibility/iphone/vision/*.

[2] American Foundation for the Blind. CCTVs/Video Magnifiers: *http://www.afb.org/prodBrowseCatResults.aspx?CatID=53*.

[3] eSight Glasses: *https://www.esighteyewear.com/technology*.

[4] Google Glass: *https://www.google.com/glass/start/*.

[5] Microsoft Hololens: *https://www.microsoft.com/hololens*.

[6] Shilkrot, R., Huber, J., Wong, M.E., Maes, P. and Nanayakkara, S. 2015. FingerReader: a wearable device to explore printed text on the go. *Proc. CHI 2015*, 2363–2372.

[7] Stearns, L., Du, R., Oh, U., Jou, C., Findlater, L., Ross, D.A. and Froehlich, J.E. 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *TACCESS*. 9, 1 (Oct. 2016), 1–38.

[8] Zhao, Y., Szpiro, S. and Azenkot, S. 2015. ForeSee: A Customizable Head-Mounted Vision Enhancement System for People with Low Vision. *Proc. ASSETS 2015*, 239–249.

TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments

LEE STEARNS, University of Maryland, College Park
URAN OH, University of Maryland, College Park; Carnegie Mellon University
LEAH FINDLATER, University of Maryland, College Park; University of Washington
JON E. FROEHLICH, University of Maryland, College Park; University of Washington

On-body interaction, which employs the user's own body as an interactive surface, offers several advantages over existing touchscreen devices: always-available control, an expanded input space, and additional proprioceptive and tactile cues that support non-visual use. While past work has explored a variety of approaches such as wearable depth cameras, bio-acoustics, and infrared reflectance (IR) sensors, these systems do not instrument the gesturing finger, do not easily support multiple body locations, and have not been evaluated with visually impaired users (our target). In this paper, we introduce *TouchCam*, a finger wearable to support location-specific, on-body interaction. TouchCam combines data from infrared sensors, inertial measurement units, and a small camera to classify body locations and gestures using supervised learning. We empirically evaluate TouchCam's performance through a series of offline experiments followed by a realtime interactive user study with 12 blind and visually impaired participants. In our offline experiments, we achieve high accuracy (>96%) at recognizing coarse-grained touch locations (*e.g.*, palm, fingers) and location-specific gestures (*e.g.*, tap on wrist, left swipe on thigh). The follow-up user study validated our real-time system and helped reveal tradeoffs between various on-body interface designs (*e.g.*, accuracy, convenience, social acceptability). Our findings also highlight challenges to robust input sensing for visually impaired users and suggest directions for the design of future on-body interaction systems.

CCS Concepts[1]: • **Human-centered computing~Gestural input** • **Human-centered computing~Mobile devices** • *Human-centered computing~Accessibility technologies* • *Computing methodologies~Machine learning algorithms*

KEYWORDS

On-body input; Accessibility; Wearable sensors; Computer Vision Applications; Blind and Low-Vision Users; Skin Texture Classification; Gesture Recognition

ACM Reference format:

Lee Stearns, Uran Oh, Leah Findlater, and Jon E. Froehlich. 2017. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* #, #, Article ## (December 2017), 23 pages.
DOI: ##.####/####

1 INTRODUCTION

On-body interaction, which employs the user's own body as an interactive surface, is an emerging area of research that offers several advantages over existing touchscreen devices, particularly for non-visual use (*e.g.*,

This work is supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.
Contact author's address: L. Stearns, Department of Computer Science, 3173 AV Williams, University of Maryland, College Park, MD 20742.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2017 Association for Computing Machinery.
2474-9567/2017/12-ARTx $15.00
DOI: 10.1145/1234

blind users). Taps, swipes, or other on-body gestures provide lightweight and always-available control (e.g., [17,21]) with an expanded input space compared to small-screen wearable devices like smartwatches (e.g., [27,31,40,42]). In addition, the proprioceptive and tactile cues afforded by on-body input can improve eyes-free interaction (e.g., [10,32,35,57]) and enable accurate input even without visual feedback compared to the smooth surface of a touchscreen [18,45]. These advantages are particularly compelling for users with visual impairments, who do not benefit from visual cues and who frequently possess a heightened sense of tactile acuity ([15,39]).

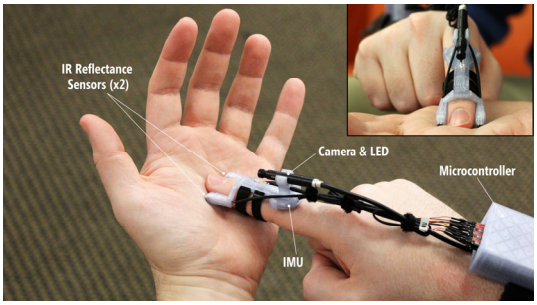


Fig. 1. *TouchCam* combines a finger-worn camera with wearable motion trackers to support location-specific, on-body interaction for users with visual impairments. See supplementary video for a demonstration.

Reliably sensing on-body input, however, is an open challenge. Researchers have explored a variety of approaches such as cameras (e.g., [4,21,50]), infrared-reflectance sensors [26,27,40], and bio-acoustics [20,28]. While promising, this prior work has not been specifically designed for or tested with visually impaired users, who likely have different needs and preferences. For example, blind users may encounter difficulty accurately aiming a camera (or other directed sensor) [25,55] and also rely more on their sense of touch [15,39] making it especially important to avoid covering the fingertips. Furthermore, prior work does not support complex gestures at multiple body locations. For example, Skinput [20] detects touches at a range of locations but not more complex gestures. In contrast, FingerPad [6] and PalmGesture [57] sense shape gestures performed on the fingertip or palm but cannot easily be extended to other locations.

Our research explores an alternative approach called *TouchCam* (Figure 1), a custom designed finger wearable to support location-specific, on-body interaction. We previously demonstrated the feasibility of recognizing body locations from small skin surface images (1–2 cm) captured using a handheld camera [53]. However, this work did not include sensor fusion, use a wearable form factor, or function in real-time. In addition, [53]'s prototype could only recognize locations (not gestures), and the images were collected under carefully controlled conditions. In this paper, we build on that work and address these limitations. TouchCam combines data from infrared reflectance (IR) sensors, inertial measurement units (IMU), and a small camera to classify body locations and gestures using supervised learning. Because TouchCam instruments the *gesturing* finger, on-body interaction is supported on a variety of locations within the user's reach while also mitigating camera framing issues. TouchCam also enables new location-specific, contextual gestures that are semantically meaningful (*e.g.*, tapping on the wrist to check the time or swiping on the thigh to control a fitness app). These features allow for flexible interface designs that can be customized based on needs of the application or user. In this paper, we explore four high-level research questions:

RQ1. How well can we recognize location-specific on-body gestures using finger- and wrist-mounted sensors?
RQ2. Which locations and gestures can be recognized most reliably using this sensing approach?
RQ3. What tradeoffs must be considered when designing and building a realtime interactive on-body input system?
RQ4. How accessible is this approach to users with visual impairments and what are their design preferences?

To address these questions, we evaluated two prototype iterations across two studies. In Study I, we demonstrate feasibility through a controlled data collection study with 24 sighted participants who performed touch-based gestures using the first iteration of our prototype (*TouchCam Offline*). In offline experiments using classifiers trained per-user, we achieve 98% accuracy in classifying coarse-grained locations (*e.g.*, palm, thigh), 84% in classifying fine-grained locations (*e.g.*, five locations on the palm), and 96% in classifying location-specific gestures. Informed by these results, we built a second prototype with updated hardware and software algorithms

to support *realtime* on-body localization and gesture recognition (*TouchCam Realtime*). In Study II, we investigate the usability and potential of the realtime system with 12 blind and visually impaired participants. Our findings validate realtime performance with our target population and highlight tradeoffs in accuracy and user preferences across different on-body inputs.

In summary, the primary contributions of this paper include: (i) two iterations of TouchCam, a novel finger-worn camera system that uses machine learning to detect and recognize on-body location-specific gestures; (ii) a quantitative evaluation of our system's accuracy and robustness across a variety of gestures and body locations; (iii) qualitative observations about the usability and utility of our on-body input approach for users with visual impairments; and (iv) design reflections for on-body gestural interfaces in terms of what locations and gestures can be most reliably recognized across users. While our prototype design is preliminary—we expect that future iterations will be much smaller and self-contained—our explorations build a foundation for robust and flexible on-body interactions that support contextual gestures at multiple body locations via supervised learning. Our primary focus is supporting users with visual impairments; however, our approach could also benefit users with situational impairments (*e.g.*, while walking or conversing) or be applied as an input mechanism for virtual reality systems (*e.g.*, for accurate touch-based input in eyes-free situations). All software code and hardware design files are open sourced and available here: https://github.com/lstearns86/touchcam.

2 RELATED WORK

Our research is informed by work on sensing on-body input, finger-worn input devices, and algorithms for texture classification and camera-based biometrics.

2.1 Sensing On-body Input

As noted in the Introduction, on-body input has several advantages over handheld or wearable touchscreen input (*e.g.*, smartphones or smartwatches) offering a larger input surface and more precise touch input even without visual cues [18,45]. However, how to sense this input and what form it should take are still open questions. Researchers have investigated a wide variety of wearable sensing approaches, including cameras [4,10,19,21,50,54,57], IR [27,40–42], ultrasonic rangefinders [30,32], bio-acoustics [20,28], magnetic fields [6], electromyography (EMG) [35], electromagnetic phase shift [62], and capacitance sensors [30,35,48,58]. These approaches support a similarly wide variety of inputs, including discrete touches at different body locations [28,35,48], continuous touch localization on the hand or wrist similar to touchscreen input [4,21,50,62], and input based on 3D finger or arm positions [4,48,50]. We summarize a subset of this prior work alongside our own in Table 1, which helps highlight the diversity of sensing approaches and on-body interaction support. While these past approaches are promising, their sensor types and placements limit the types of interactions that they can support.

First, the interaction space is often constrained to a small surface (*e.g.*, wrist or arm) or to a narrow window in front of the user. Approaches using cameras mounted on the upper body (*e.g.*, [10,19,21,54]) restrict interactions to a pre-defined region within the camera's field of view. *OmniTouch* [21], for example, can only detect gestures on the hands or arms in a relatively small space in front of the user. Similarly, approaches using sensors mounted on one wrist or hand to detect gestures performed by the other hand (*e.g.*, [4,20,27,28,30,41,50,57,58,62]) limit on-body interactions to a relatively small area around the sensors. Some approaches such as *Touché* [48] or *iSkin* [58] are more flexible but still require instrumentation at the target interaction location, which limits scalability. In contrast, our approach places sensors on the *gesturing* finger, supporting input at a variety of body locations within the user's reach without requiring additional instrumentation. Further, while not evaluated in this paper, our system could be readily extended to interact with surfaces beyond the body.

Second, prior work attempts to either identify touched body locations or detect motion gestures but not both. For example, *Touché* [48] and *Botential* [35] can localize touch input at various locations on the body using EMG or capacitance sensors. However, these systems cannot recognize relative surface gestures such as directional swipes. In contrast, systems such as *PalmGesture* [57], *SkinTrack* [41], or *WatchSense* [50] can estimate precise 2D touch coordinates, enabling complex gesture interactions like shapes. However, these methods require sensors affixed on or near the interaction surface to achieve such precision, and they therefore cannot easily be extended to recognize multiple locations. Our approach uses a small camera to identify touched locations, augmented by inertial and IR sensors for robust gesture recognition; together, these sensors enable location-specific gestures.

2.2 Finger-worn Input Devices

TouchCam is worn on the user's gesturing finger and is informed by the growing space of finger-mounted wearables (see [49] for a review). These devices come in many different form-factors (*e.g.*, rings [2,38,43,51], nails [6], below the finger pad [61], or between the fingers [4]), and are used for many different purposes (*e.g.*, reading [51], object/face identification [4,38], or gestural input [2,4,26,43]). For example, *FingerPad* [6] uses a paired magnet and magnetometer on the back of the user's thumb and finger to enable subtle 2D gestural input, while *Light Ring* [26] uses a small gyroscope and IR sensor embedded in a ring to enable input on a variety of surfaces. We focus on touch-based input rather than midair gestures (*e.g.*, [5]) because of the speed and accuracy advantages for non-visual use [45]. We also position our sensors above the gesturing finger to avoid interfering with the user's touch sensitivity and to support flexible input at many different locations.

Most closely related to our work are systems that include finger-worn cameras. *EyeRing* [38], for example, uses a camera ring to identify physical objects in the surrounding environment while *HandSight* [13,51,52], our own prior work, uses a tiny camera atop the finger to assist users with visual impairments in reading and exploring printed text. *CyclopsRing* [4] uses a fisheye lens camera clipped between the fingers to detect one-handed gestures or touch gestures from the user's other hand; however, as discussed above, it cannot easily extend to other locations to support location-specific gestures. *Magic Finger* [61], which partially inspired our approach, uses a tiny camera and an optical mouse sensor worn below the pad of the finger to support touch input on almost any surface, as well as texture-based identification of that surface. However, only two body locations were evaluated and location-specific gestures were not investigated.

Building on the above, we investigate a different set of sensors (optical and inertial), broaden the on-body interaction possibilities (*e.g.*, location-specific gestures), and evaluate a larger number of on-body locations (15

System Name	Sensor type	Sensor placement	On-body Interaction Space	Interaction type
OmniTouch [21]	Camera (Depth)	On the shoulder	On or above the hands or arms (limited by camera FoV)	Continuous touch locations
Touché [48]	Capacitive	Flexible (one on wrist, one elsewhere on body)	Flexible (requires the target location to be instrumented)	Discrete touch locations, body or hand pose
CyclopsRing [4]	Camera (RGB, Fisheye Lens)	Between fingers of passive hand (for on-body input)	On or above the instrumented hand	Continuous touch locations, touch gestures, hand pose
Botential [35]	EMG, capacitive	On the wrist (or arm, leg)	Flexible, different body parts	Discrete touch locations
ViBand [28]	Bio-acoustic	On the wrist	On the instrumented hand or arm	Discrete touch locations, non-directional gestures
SkinTrack [62]	Electromagnetic phase shift	On the wrist, ring on opposite hand	On the skin surface around the instrumented wrist	Continuous touch locations, touch gestures
WatchSense [50]	Camera (Depth)	On the wrist, facing toward fingers	On or above the instrumented hand (limited by camera FoV)	Continuous touch locations, touch and midair gestures
TouchCam (Our Work)	Camera (Grayscale), IMU, IR	On top/side of the gesturing finger and wrist	Flexible (does not require additional instrumentation)	Discrete touch locations, touch gestures

Table 1. Overview of several recent on-body input approaches alongside our own work.

vs. 2 in Magic Finger, for example). We also evaluate our approach with visually impaired participants to explore their specific needs and behaviors, which may differ from sighted users.

2.3 Texture Classification and Camera-Based Biometrics

Work in biometrics has demonstrated that the skin of the hand—specifically, the palm and fingers—contains many distinctive visual features that can be used to identify individuals [7,36]. While skin features are commonly used for identification, they can also support on-body localization—as demonstrated by our initial work [53]. In [53], we extended biometrics algorithms, including those for partial fingerprint and palmprint recognition [12,24] as well as techniques that use webcam or mobile phone camera images for identification and verification [7,9,37,60]. We classified images of small patches of skin on the hand and wrist (patch size ~1-2 cm^2) captured with a handheld camera and achieved high accuracy (>96% across 17 body locations). Our algorithms relied primarily on a texture classification approach inspired by Magic Finger [61] with a local binary pattern (LBP) texture representation and a support vector machine (SVM) classifier. However, as noted in the introduction, our prior work was not wearable, did not function in real-time, only supported localization and not gesture recognition, the data was collected under carefully controlled conditions, and no user studies were conducted. This paper addresses each of those limitations.

3 TOUCHCAM OFFLINE: INITIAL WEARABLE PROTOTYPE

We describe our first prototype, TouchCam Offline, which we evaluate offline using data collected from a controlled study. Study I focuses on addressing RQ1 and RQ2: *how accurately can we recognize location-specific on-body gestures* and *which locations and gestures can be recognized most reliably?* Our results inform the development of a realtime prototype, which is evaluated in Study II (Section 5).

3.1 Prototype Hardware

The TouchCam Offline hardware consists of: (i) a finger-worn multi-sensor package that includes two infrared (IR) reflectance sensors, an inertial measurement unit (IMU), and a small camera with an adjustable LED for illumination; and (ii) a wrist-worn microcontroller with a second IMU, which simulates a smartwatch and provides additional sensing. The finger-based sensors are mounted on three laser-cut rings and positioned to avoid impeding the user's sense of touch, which is particularly important for users with visual impairments. See Figure 2a.

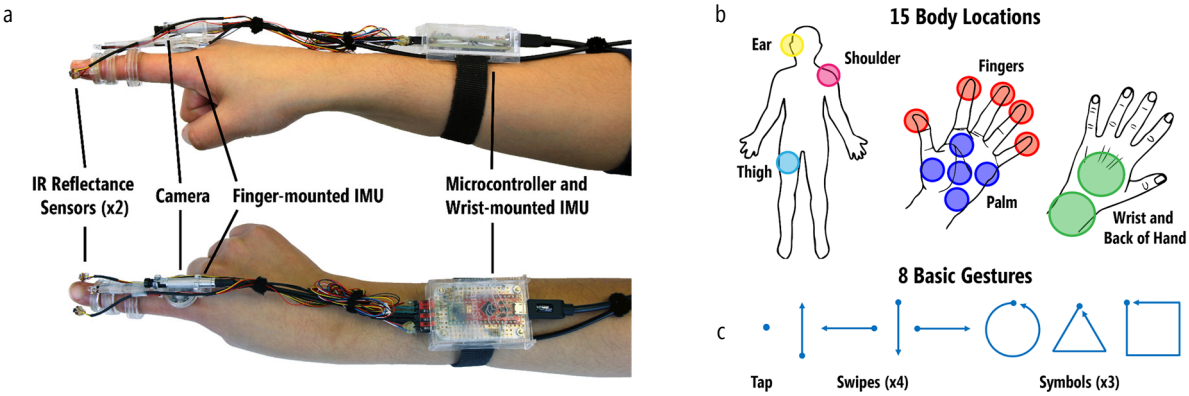


Fig. 2. (a) TouchCam Offline showing the finger and wrist-worn sensors and microcontroller. (b) Fifteen fine-grained body locations (individual circles) within six coarse-grained locations (denoted by color), and (c) eight basic gestures.

Infrared Reflectance Sensors. The two IR sensors[2] (each 2.9 mm × 3.6 mm × 1.7 mm) have a sensing range of ~2–10 mm and are used to detect touch events and to aid in recognizing gestures. Unlike Magic Finger [61], which places optical sensors directly on the pad of the finger, we mount the sensors on the sides of the front-most ring, approximately 5 mm from the fingertip to avoid interfering with tactile sensitivity.

Camera Sensor. A small (6 mm diameter) CMOS camera[3] is mounted atop the user's index finger, providing 640 × 640 px images at up to 90 fps with a 30° diagonal field of view (FOV). We use grayscale images from the camera to classify the touch location, extracting both texture and 2D point features. We also estimate visual motion between video frames to assist in classifying gestures. The camera includes a manually adjustable lens with a focal distance that varies from 15 mm to ∞. Because distances shorter than 30–40 mm provide a very narrow range of focus, we positioned the front of the lens ~50 mm from the user's fingertip. This setup provides an effective FOV of ~27mm across the diagonal when the finger is touching a surface. To prevent lateral rotation and to fix the FOV center near the touch location, the camera is attached to two rings 15 mm apart. A bright LED (3 mm diameter, 6000 mcd, 45° angle) mounted below the camera lens illuminates the touch surface regardless of ambient lighting.

Inertial Measurement Units. Two IMUs[4] are mounted on the user's hand: one below the camera on the index finger and one on the wrist. We include two IMUs to examine the effect of sensor location on classification performance. The IMUs provide motion information at ~190 Hz, and each contains a three-axis accelerometer, gyroscope, and magnetometer. While the camera offers rich contextual information about a scene, its field of view and frame rate limit performance during quick motions. Therefore, the IMUs are our primary sensor for detecting motion and classifying gestures. The orientation of the gesturing finger and/or wrist may also be useful for distinguishing body locations (*e.g.*, ear vs. thigh) although this is posture dependent. The IMUs are calibrated to correct magnetic bias and to establish a stable orientation estimate (described in Section 3.2). The calibration process consists of rotating the unit along each axis for a few seconds and is performed only once per study session—although future explorations may require repeated calibration to ensure long-term stability.

Sensor Placement and Microcontroller. We designed custom laser-cut rings in multiple sizes (13–24 mm inner diameter in 0.5mm increments) with detachable sensors to fit each user. As shown in Figure 2a, the rings are worn on the index finger near the first and second joints. The IR and IMU sensors are controlled via a microcontroller[5] mounted on a Velcro wristband, and the camera and microcontroller are connected to a desktop computer[6] via USB cables. All data is logged, timestamped, and analyzed *post hoc* on a desktop.

3.2 Input Recognition Algorithms

To recognize localized on-body input, we developed a four-stage approach: (i) touch segmentation; (ii) feature extraction; (iii) location classification; (iv) gesture classification. The two classification stages—location and gesture—are trained individually for each user and combine readings from multiple sensors for robustness. While the algorithms described next could be trained on any arbitrary set of locations and gestures, in our study, we evaluated six coarse-grained body locations (*fingers*, *palm*, *back of hand or wrist*, *ear*, *shoulder*, and *thigh)* with 15 fine-grained locations (*thumb/index/middle/ring/pinky finger*, *palm up/down/left/right/center*, *back of hand*, *outer wrist*, *ear*, *shoulder*, and *thigh*) and 8 basic gestures (*tap*, *swipe up/down/left/right*, *circle*, *triangle*, and *square*)—see Figures 2b and 2c. These locations are visually distinctive and can be located easily even without sight while the gestures are simple shapes that can be drawn with a single touch down/up event.

[2] Fairchild Semiconductor QRE113GR

[3] Awaiba NanEye GS Idule Demo Kit

[4] Adafruit Flora LSM9DS0

[5] Sparkfun Arduino Pro Micro (5V/16MHz)

[6] Dell Precision Workstation, dual Intel Xeon CPU @2.1GHz, NVIDIA GeForce GTX 750Ti

Stage I: Touch Segmentation. Our input recognition algorithms receive a sensor stream consisting of video, IMU, and IR data. We segment this input stream by detecting touch-down and touch-up events using the IR sensor readings, which represent distance from the touch surface (lower values are closer). While for real-world use, a segmentation approach would need to identify these touch events within a continuous stream of data, to evaluate this initial prototype we made several assumptions to simplify the process (we eliminated these assumptions for the realtime prototype, described later). Based on experiments with pilot data, we developed a straightforward threshold-based approach using a variable threshold that was set to 90% of the maximum IR value observed across the input stream for each trial. Within a trial, a *touch-down* event is triggered when either of the two IR values crosses below the threshold, while a *touch-up* event is triggered when both cross above the threshold. To be conservative, we assume that each trial contains a single gesture and segment the entire gesture from the first *touch-down* event in the trial to the last *touch-up* event. We crop each input stream to include only the sensor readings and video frames that occurred between the touch-down and touch-up event timestamps.

Stage II: Feature Extraction. In Stage II, we extract static orientation and visual features for localization, and motion features for gesture classification. We describe each in turn below (see Table 4 in the Appendix for more details).

Localization Features. To extract static features for localization, we first determine the video frame that has the maximum focus in the segmented sequence, since it is the most likely to contain recognizable visual features. We define focus as the total number of pixels extracted using a Canny edge detector [3] tuned with a small aperture ($\sigma = 3$) and relatively low thresholds ($T_1 = 100, T_2 = 50$) to detect fine details. While this approach does not account for all image quality problems—motion blur in particular can cause it to fail—it is highly efficient and, in general, detects a much greater number of edges for images that are in focus than for those that are not. We verified this trend empirically using pilot data. We then extract several features for the selected video frame, which include: (i) raw IR sensor readings, (ii) the estimated IMU orientation, (iii) image texture features for coarse-grained classification, and (iv) 2D image keypoints for geometric verification to distinguish between locations with similar textures (*i.e.*, fingertips, palm locations, back of wrist or hand).

The orientation of each IMU is estimated by applying a Madgwick filter [33] on a sequence of raw accelerometer, magnetometer, and gyroscope readings resulting in a 4D orientation vector (*i.e.*, quaternion). The filter is a standard sequential optimization approach to estimating IMU orientations that is updated at each time step. Our initial calibration procedure includes briefly rotating the device in all directions so that the filter can converge to an accurate orientation estimate. The orientation estimate at the selected video frame is used as a 4-dimensional feature vector (*W*, *X*, *Y*, and *Z*) for each IMU and concatenated into an 8-dimensional vector when both IMUs are used.

The image-based features are extracted similarly to our prior work [53]: To represent image texture, we use a variant of local binary patterns (LBP) that is robust to changes in illumination and that achieves rotation invariance while exploiting the complementary nature of local spatial patterns and contrast information [16]. While we explored other common texture-based methods such as Gabor histograms [56] and wavelet principal components [11], we found that they offered negligible improvements over LBP despite their increased computational complexity. We extract uniform LBP patterns and local variance estimates from an image pyramid with eight scales to capture both fine and coarse texture information. Specifically, we use $LBP^{riu2}_{12,2}/VAR_{12,2}$ with 14 uniform pattern bins and 16 variance bins as defined in [16]. These values are accumulated into a histogram with 224 bins for each scale, all concatenated to obtain a 1792-element feature vector. To resolve ambiguities and ensure geometric consistency, we extract custom keypoints at locations with a high Gabor filter response at two or more orientations, which tend to lie at the intersections of ridgelines or creases. This approach was inspired by [23]. We use the Gabor energy in a 16 × 16px neighborhood around the keypoint as a descriptor extracted at multiple orientations to ensure rotation invariance. See [53] for full details.

Motion Features. For gesture classification, we extract motion features from the sensor readings within the segmented timeframe (these are treated independently from the localization features). We use three standard signal preprocessing steps on the raw IMU and IR sensor readings: smoothing, normalization, and resampling. We first smooth the raw values using a Gaussian filter (σ =13, optimized based on pilot data) to reduce the effect of sensor noise and then normalize the smoothed sequence by subtracting its mean and dividing by its standard deviation. To obtain a fixed length sequence for robustness to variations in speed, we resample the sensor readings using linear interpolation at 50 equally spaced discrete time steps. These values, however, are still sensitive to small variations in speed and orientation. Thus, similar to [59], for each IMU and IR sensor we compute summary statistics for windows of 20 samples at 10-step increments (*i.e.,* four windows): mean, minimum, maximum, median, and absolute mean. Finally, for the 50 resampled accelerometer, magnetometer, and gyroscope readings, we compute *x-y*, *x-z*, and *y-z* correlations. The result is 639 features for each IMU and 70 for each IR sensor, which we concatenate into a single feature vector to use when classifying gestures.

We also extract motion features from the video frames between the touch-down and touch-up events. Because we support touch-based gestures only on flat or nearly flat surfaces, it is sufficient to estimate a global 2D motion vector for each frame; we do so using a template-matching approach. First, we down-sample each image from 640 × 640px to 160 × 160px resolution for efficiency and noise robustness. Next, for each frame we extract a 40 × 40px region centered within the previous frame to use as a template, which we then match against the current frame using a sliding window to compute the normalized cross-correlation [29]. The position of the pixel with the highest cross-correlation value identifies the most likely displacement between frames, yielding a 2D motion vector estimate. Because images with higher contrast are more likely to yield reliable motion estimates, we weight each motion vector by an estimate of the frame's contrast (the image variance). As with the other motion features, we smooth the motion estimates by applying a moving average (window size = 10). We then re-sample 50 points from this sequence of motion vectors and compute summary statistics as with the IMU and IR sensor readings to obtain a fixed-length vector of 140 features for use in gesture classification.

Stage III: Localization. Once we have extracted localization and motion features, we begin independently classifying on-body locations (Stage III) and gestures (Stage IV). For localization, we rely primarily on static visual features from the camera with IMU orientations and IR reflectance values to resolve ambiguities.

Our image-based touch localization algorithms function identically to our prior work [53]. We use a two-level location classification hierarchy: first classifying the location as one of the six course-grained regions then refining that location estimate where possible to finer-grained regions. In our offline user study (Study I), coarse-grained regions include *fingers*, *palm*, *back of hand or wrist*, *ear*, *shoulder*, and *thigh* while fine-grained regions include specific fingertips, locations on the palm, and on the back of hand versus the wrist (Figure 2b). Some coarse-grained locations are not subdivided at this second level due to a lack of distinctive features—in the case of our study, the *ear*, *shoulder*, and *thigh* are not subdivided. We first classify the texture features into a coarse-grained location using an SVM[7] then perform template matching against only the stored templates from that location to estimate the fine-grained location. Finally, we perform geometric verification using the extracted 2D point features to ensure a correct match.

At both levels of the classification hierarchy, we resolve ambiguities using a sensor fusion approach. We combine predictions based on the static visual features from a video frame with predictions based on the IMU orientation and IR reflectance features with the same timestamp as that frame. Since the scales, lengths, and types of these feature vectors are all very different, rather than concatenating the features for use with a single classifier we instead train a separate SVM with a Gaussian kernel on the non-visual features. To robustly combine the predictions from the two disparate localization classifiers (one for the camera features and one for the IR and IMU features), we first tune the SVMs to output normalized probability predictions for each class

[7] Aforge.NET: http://www.aforgenet.com (used for all SVM and neural network classifiers)

using Platt scaling, as is standard [47]. We concatenate these predictions into a single feature vector, which we then use to train a third sensor fusion classifier that automatically learns how to prioritize sensors based on prediction confidence and location class. Inspired by [8], we use a feedforward neural network for this sensor fusion classifier. Our network has one fully connected hidden layer for flexibility of functional representation, and a softmax output layer for multiclass output; it is trained using resilient backpropagation [34]. The final output of our classification process is a combined location prediction from among the six coarse-grained and fifteen fine-grained classes with approximate likelihoods for each class (sorted from most to least likely).

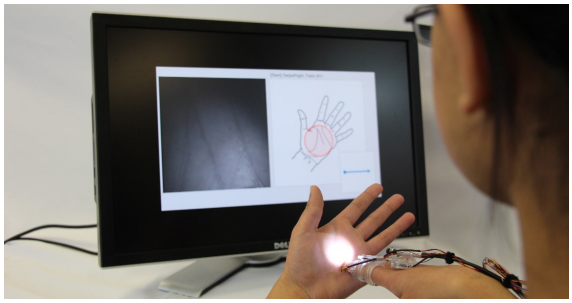
Stage IV: Gesture Classification. Gesture classification is performed independently of localization using an additional SVM. As with texture, SVM classifiers are commonly used for classifying gesture features because they are robust and efficient for problems with high dimensionality. We use a linear kernel with feature weights that were optimized for performance across all participants. For the evaluation presented in Section 3.4, we trained an SVM to classify the following gestures as shown in Figure 2c: *tap*, *swipe up*, *swipe down*, *swipe left*, *swipe right*, *circle*, *triangle*, and *square*.

3.3 Study I: Data Collection and Dataset For Offline Experiments

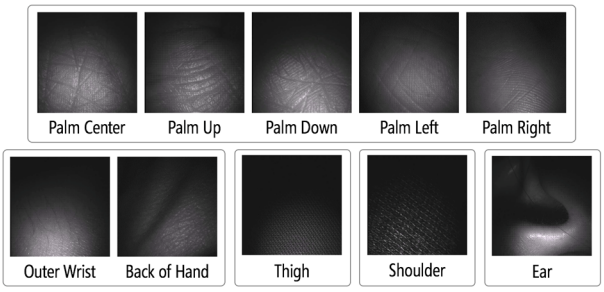
To evaluate our initial prototype and algorithms, we performed offline experiments using data collected from twenty-four participants. Each participant performed a series of location-specific on-body input tasks with our hardware prototype. We were specifically interested in investigating our first two research questions enumerated in the Introduction: (i) How accurately can we recognize location-specific on-body gestures with a finger-worn camera and auxiliary sensors (IMU, IR)? (ii) Which body locations and gestures can be recognized most reliably using our approach?

Participants. Twenty-four right-handed participants (16 female) were recruited via campus e-mail lists and word of mouth. Their average age was 28.9 (*SD*=7.95, *range*=19–51). All participants had normal vision as the goal of this study was to assess our algorithms and not issues related to usability or accessibility. Participants were compensated $25 for their time.

Data Collection Apparatus. During data collection, participants wore the TouchCam Offline prototype. As described in Section 3.1, we selected ring sizes to fit the participant's finger and adjusted positioning to ensure a consistent sensor range. A custom application written in C# displayed visual task prompts and a live feed from the finger-worn camera to assist with framing the target locations (Figure 3a). All IMU and IR sensor readings and camera video frames were logged with timestamps along with ground-truth touch location and gesture labels for each trial.



(a) Participant following on-screen data collection protocol



(b) Example skin images from Study I

Fig. 3. (a) Data collection setup showing our prototype, location and gesture instructions, and camera video feed. (b) Example skin-surface images recorded by our finger-mounted camera (fingerprint images omitted to protect our participants' privacy).

Procedure. The procedure lasted up to 90 minutes. After a brief demographic questionnaire and setup period (*i.e.*, selecting rings, putting on the prototype), participants completed the following tasks, in order:

1. *Location-specific touches.* Participants touched and held their finger in place at 15 locations (Figure 2b) with each location prompted visually on a monitor (Figure 3a). After confirming the location and image quality, the experimenter logged the current location (*e.g.,* timestamp, location label) and triggered the start of the next trial. Participants completed 10 blocks of trials, where each block consisted of a different random permutation of the 15 locations (150 trials in total). In total, this dataset includes 3600 *location-specific touches* across all participants. Example images are shown in Figure 3b.
2. *Location-specific gestures.* Participants performed the eight basic gestures: *tap*, *swipe up*, *swipe down*, *swipe left*, *swipe right*, *circle*, *triangle*, and *square* (Figure 2c) at three body locations: the *palm*, *wrist*, and *thigh*. These locations were selected from the 15 locations in the first task because they are easy to access, unobtrusive, and have a relatively large input area thus allowing for more complex gestures. As with the first task, participants completed 10 blocks of trials, where each block consisted of a different random permutation of the 24 gesture and location combinations (240 trials in total). This dataset includes 5,760 *location-specific gestures* across all participants.

3.4 Study I: Offline Experiments and Results

To investigate the accuracy of our location and gesture classification algorithms, we performed a series of offline experiments using the gathered data. Below, we evaluate coarse-grained localization, fine-grained localization, and location-specific gesture classification as well as the effect of each sensor on performance (*e.g.,* finger-worn *vs.* wrist-worn IMUs). We compare sensor combinations using paired t-tests and Holm-Bonferroni adjustments to protect against Type I error [22].

Training and Cross Validation. All of our experiments use leave-one-out cross validation and train and test on a single user's data. Specifically, each experiment uses all available data from a single participant for training the location and gesture classification SVMs with a single sample set aside for testing. The localization and gesture classifiers are trained independently. The experiment is repeated for each sample and averaged across all possible combinations.

Touch Localization. To examine the accuracy of our on-body localization algorithms, we used the location-specific touch dataset. Since our localization approach is hierarchical, we analyze performance at both the coarse-grained level (6 classes) and the fine-grained level (15 classes).

We first report primary localization results using all available sensor readings (*i.e.,* sensor fusion results). At the coarse-grained level, we achieve 98.0% (*SD*=2.3%) average accuracy. This is reduced to 88.7% (*SD*=7.0%) at the fine-grained level. Table 2 shows the accuracy breakdown by class. The worst performing coarse-grained classes were the wrist/hand and ear, both at 93.8%, possibly due to their highly variable appearance and fewer distinctive visual features. In contrast, the fingers and palm perform best at 99.6% and 99.1% respectively although the individual fine-grained classification accuracies were lower. These results suggest that care must be taken in selecting body locations that are both visually distinctive and easy for participants to return to repeatedly. A qualitative analysis of our dataset revealed issues that account for some of the error: approximately 5% of the images gathered had focus, contrast, or illumination issues that interfered with

	Palm	Fingers	Wrist/Hand	Ear	Shoulder	Thigh
Palm	99.1%	0.5%	0.4%			0.1%
Fingers	0.3%	99.6%				0.1%
Wrist/Hand	5.0%	0.2%	93.8%	0.2%	0.6%	0.2%
Ear	4.2%	0.4%	1.2%	93.8%	0.4%	
Shoulder	0.8%		0.4%		98.8%	
Thigh	2.3%		0.4%			97.3%

(a) Coarse-grained Accuracy

	Up	Down	Left	Right	Center
Palm	84.6%	78.5%	85.0%	83.1%	91.5%

	Thumb	Index	Middle	Ring	Pinky
Fingers	93.1%	85.4%	81.5%	88.1%	91.9%

Outer Wrist	Back of Hand
87.3%	88.8%

Ear
93.8%

Shoulder
98.8%

Thigh
97.3%

(b) Fine-grained Accuracy

Table 2. Classification percentages averaged across 10 trials and 24 participants. (a) Accuracy for the six coarse-grained classes. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row); empty cells indicate 0%. (b) Accuracy for the 15 fine-grained classes, grouped by corresponding coarse-grained class.



Fig. 4. Approximately 5% of the images we collected had poor focus, contrast, or illumination, preventing robust feature extraction. We adjusted the camera and LED to mitigate these issues for TouchCam Realtime.

extracting recognizable image features; see Figure 4. We took steps to mitigate these problems for the next iteration of TouchCam.

To investigate the effect of each sensor on localization performance, we repeated the classification experiment with the sensors individually and in combination. As expected, the camera is by far the most accurate single sensor for classifying location, with a coarse-grained accuracy of 97.5% (*SD*=2.6%) followed by the finger-based IMU at 75.6% (*SD*=11.6%). Notably, the camera is significantly better even compared to the 87.5% (*SD*=7.0%) accuracy of combining all other sensors ($p<0.001$, $t_{23}=7.12$, $d=1.92$). No significant differences were found between the camera alone or combined with other sensors, which suggests that the camera alone is sufficient for course-grained classification. At the fine-grained level, the camera is again the most accurate sensor (84.0%) even compared to all other sensors in combination (52.9% accuracy; *SD*=12.0; $p<0.001$, $t_{23}=16.74$, $d=2.99$). But, unlike at the coarse-grained level, adding any of the other three sensors to the camera further increases accuracy, with the highest accuracy (88.7%) resulting from the combination of all available sensors.

Location-Specific Gesture Classification. To explore the possibility of supporting location-specific gestures, we conducted a classification experiment with the data from the location-specific gesture task (24 classes: 3 locations × 8 gestures). First, we classified the location using the image features from the camera (extracted from the video frame with maximal focus as described above). Since the location features from the IR and IMU sensors did not make a significant difference at the coarse-grained level, we omitted them here. Location accuracy for these three locations was 99.1% (*SD*=1.0%). Next, we classified the gesture using the motion features from all of the sensors (IMU, IR, and camera) achieving an accuracy of 96.6% (*SD*=2.6%). Finally, we combined the classification predictions and calculated the overall location-specific gesture classification accuracy across all 24 classes, which was 95.7% (*SD*=3.2%).

As a secondary analysis, we again examined classification accuracy as a function of each sensor (Figure 5) but this time for the 24 location-specific gestures. In general, adding more sensors significantly improves classification accuracy, although as a practical matter the differences between the pair of IMUs and other more complex combinations are fairly small (see the Appendix for statistical comparisons).

Efficiency. For our initial prototype and algorithm development, our primary aim was to investigate the feasibility and accuracy of our approach rather than develop a realtime system. As such, our TouchCam Offline algorithms are slow. On our desktop computer (the Dell Precision Workstation described in Section 3.1), the

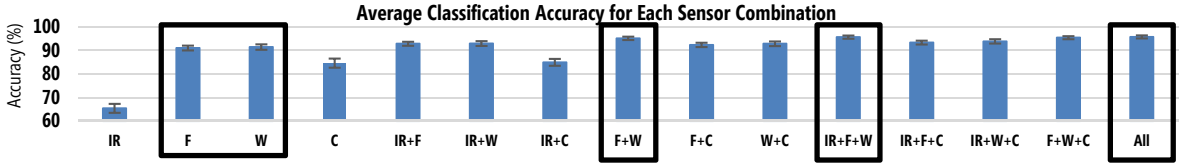


Fig. 5. Mean classification accuracy using different sensor combinations to classify location-specific gestures. Boxes indicate the best sensor combinations as additional sensors are added, with each box significantly outperforming the last (from left to right). There was no significant difference between the finger- and wrist-mounted IMUs.

image feature extraction and localization stages required, on average, two seconds *per frame* to process and classify an image. The most computationally demanding stage was the geometric verification process, which required approximately 243,000 feature comparisons on average. The other stages' computation times are comparatively negligible.

3.5 Summary of Study I Findings

Our results address our first two research questions demonstrating the feasibility of recognizing location-specific gestures using finger- and wrist-worn sensors. While our experiments show advantages with sensor fusion when classifying both location and gesture, the practical differences are relatively small suggesting that we can simplify our algorithms by using each sensor type for the task for which it is best suited (*i.e.*, IR sensors for touch detection, camera for localization, and IMUs for gesture recognition). Individual accuracies per location suggest limits to the localization granularity of our algorithms, which performed well (≥98%) for coarse-grained locations but were less accurate (89%) for fine-grained locations. These results could likely be improved with better camera hardware (*e.g.*, higher resolution, autofocus) and with more complex finger/palm print recognition algorithms. However, the high accuracy during our location-specific gesture experiment (96%) suggests that such steps may not be necessary for us to begin investigating these interactions with visually impaired users. We built upon these findings to implement the next iteration of TouchCam, described below.

4 TOUCHCAM REALTIME: IMPROVED INTERACTIVE PROTOTYPE

Based on our Study I findings, we designed *TouchCam Realtime*, a realtime version of our offline system with updated hardware and algorithms. We first describe key changes to improve robustness and enable realtime interactions (addressing RQ3) before validating the new classification algorithms using the Study I data.

4.1 Realtime Prototype Hardware

TouchCam Realtime's hardware (Figure 6) embeds all finger-mounted components in a single unit, which is attached to the user's finger using a pair of Velcro strips to allow greater freedom of motion than the rigid rings from the previous version. This updated design is more stable and durable. The camera and IR sensors are repositioned to capture more consistent images and improve the reliability of touch detection, respectively. Although Study I found an accuracy advantage when using two IMUs (~4%), we decided to remove the wrist-mounted IMU to simplify our hardware and algorithms. We compensated for the potential drop in accuracy by doubling the remaining (finger-mounted) IMU's sampling rate. This change reduced the number of features used to classify gestures and the number of examples needed for training.

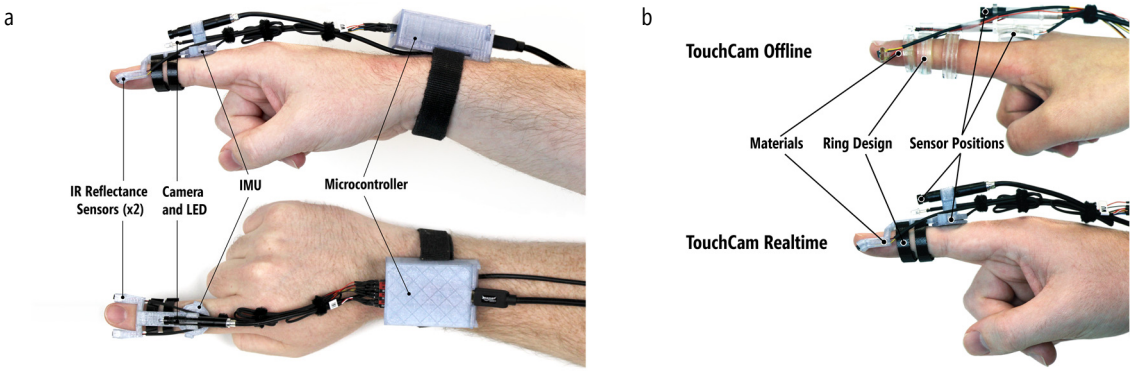


Fig. 6. (a) TouchCam Realtime prototype showing the finger and wrist-worn sensors and wrist-worn microcontroller. (b) Comparison between TouchCam Offline and Realtime hardware.

4.2 Realtime Input Recognition Algorithms

We made several changes to our input recognition algorithms to support realtime operation. First, we optimized our localization algorithms to run in realtime on a GPU and removed the computationally costly geometric verification step. Second, we updated the touch detection stage to support continuous use. Finally, we improved the gesture recognition stage, making it more robust to changes in orientation and pose.

Localization Algorithm Changes. As noted previously, our offline localization algorithms required up to two seconds per frame, primarily limited by geometric keypoint matching between image templates. Simply removing keypoint matching increased our frame rate from 0.5fps to ~18.5fps, but with a ~9% reduction in Study I's fine-grained localization accuracy. To address this loss, we made three updates to our localization algorithms. First, we used an alternate LBP algorithm that better preserved spatial features [64], which increased the number of texture features per image from 1792 to 15,552. Second, we averaged class probability predictions across 20 video frames, a number selected after pilot tests to balance accuracy and latency. And third, we reduced the number of fine-grained locations, omitting the five fingertip locations evaluated in Study I. This decision was not solely due to algorithmic performance—the fingertips proved difficult for participants to capture reliably even with visual feedback due to the sensors' positioning and small field of view. Also, while the fingertips are convenient locations for static touch-based input, they are too small to easily support gestural input. Finally, we implemented parallel GPU versions of our algorithms, which further improved the localization speed to 35.7 fps.

Touch Detection Algorithm Changes. To improve robustness and support continuous use, we made minor changes to the touch detection algorithms. We applied a moving average filter to the IR values to reduce sensor noise (*window size* = 50ms), and triggered *touch-down* and *touch-up* events when the sensors crossed a fixed threshold that was the same across all users rather than derived per gesture as with the offline system. This threshold was fixed at 90% of the maximum possible value the sensor could register, which we determined empirically to be robust to changes in ambient lighting and to work well for skin and clothing surfaces. To ensure that we captured the full gesture (and to support the double-tap gesture), we placed a delay of 100ms on the *touch-up* event and canceled it if the user touched down again within that period.

Gesture Recognition Algorithm Changes. Lastly, we made improvements to the gesture recognition algorithms. To compensate for variations in orientation and pose when performing gestures, we first rotated the IMU sensor readings relative to the estimated orientation at the start of the gesture (the *touch-down* event). We discarded the magnetometer readings after this step since they were still overly sensitive to orientation and location. These changes allowed us to build a pre-trained cross-user gesture classifier with 1,720 samples in place of the individual classifiers used in Study I.

4.3 Validation of Realtime Algorithms

To test our updated algorithms and establish a performance benchmark for our realtime system, we conducted classification experiments on the data gathered during Study I. The average 10-fold cross-validation accuracy on the location-specific touches dataset was 97.5% (*SD*=2.4%) at the coarse-grained level (6 classes) and 84.5%

	Palm	Fingers	Wrist/Hand	Ear	Shoulder	Thigh
Palm	98.5%	0.8%	0.5%	0.1%		0.1%
Fingers	0.3%	99.7%				
Wrist/Hand	4.0%	0.4%	95.4%			0.2%
Ear	5.4%	2.1%		92.1%	0.4%	
Shoulder	1.7%	0.4%	2.1%		95.4%	0.4%
Thigh	1.7%		2.1%	0.4%		95.8%

(a) Coarse-grained Accuracy

	Up	Down	Left	Right	Center
Palm	83.8%	82.5%	80.8%	85.4%	89.6%

	Thumb	Index	Middle	Ring	Pinky
Fingers	92.1%	71.3%	71.3%	73.8%	79.6%

Outer Wrist	Back of Hand	Ear	Shoulder	Thigh
87.5%	85.8%	92.1%	95.4%	95.8%

(b) Fine-grained Accuracy

Table 3. TouchCam Realtime performance on Study I dataset. (a) Coarse-grained classification percentages, averaged across 10 trials and 24 participants. Each cell indicates the percentage of images assigned to a predicted class (column) for each actual class (row). (b) Fine-grained classification percentages, averaged across the corresponding coarse-grained classes.

(*SD*=8.2%) at the fine-grained level (15 classes), which is nearly identical to the TouchCam Offline system—see Table 3. The five finger locations were most impacted by the removal of the geometric verification step. Localization accuracy on the location-specific gestures dataset remains similarly high at 98.6%. As mentioned above, efficiency increased considerably: from 0.5fps to 35.7fps (a ~70x speedup).

5 STUDY II: REALTIME EVALUATION WITH VISUALLY IMPAIRED PARTICIPANTS

To assess the performance and accessibility of TouchCam Realtime under more realistic conditions and with our target population (RQ4), we conducted a second study. We recruited 12 blind and visually impaired participants who performed common interactions with TouchCam such as checking the time or reading text messages. We focus primarily on issues impacting the accuracy and usability of our system (see [46] for more about the interaction designs and participant feedback).

5.1 Study II: Method

Participants completed an adaptive calibration procedure for training and then used TouchCam Realtime to perform tasks using three on-body interaction techniques.

Participants. Twelve participants (7 female, 5 male) were recruited through email lists, local organizations for people with visual impairments, and word of mouth. Nine participants were blind and three had low vision. The average age was 46.2 years old (*SD*=12.0, *range*=29–65). All participants were smartphone users (11 iPhone, 1 Android) and all reported using a screenreader either "all" or "most" of the time. Participants were compensated $60 for time and travel.

Apparatus. Throughout the study, participants wore the TouchCam Realtime prototype on their dominant hand. We assisted participants with putting on the ring and wristband and adjusted positioning to ensure consistent sensor readings. A custom C# application controlled a semi-automated training process, provided audio and synthesized speech cues during the tasks, and displayed a camera and sensor view for the researcher to ensure correct positioning. All sensor readings and video frames were logged with timestamps.

Location and Gestures. Based on observations made during Study I, we refined the locations and gestures for Study II. We reduced the coarse-grain set from 6 to 4 locations and the fine-grain set from 15 to 9 locations. Specifically, we replaced the back-of-the-hand location with the inner wrist due to inter-class similarity with the outer wrist, removed the shoulder location for ergonomic reasons, and removed the five finger locations because without 2D keypoint matching and geometric verification, classification accuracy for this region was considerably lower. The updated set of locations included: the palm (up, down, left, right, and center), the wrist (inner and outer), the thigh, and the ear (Figure 7).

While Study I showed that TouchCam can support a variety of touch-based gestures, for Study II we specifically modeled our interactions after Apple's VoiceOver[8] and Google's TalkBack[9]—two popular gesture-based mobile screenreaders for non-visual use. In total, we support 6 gestures, including: *swipe left* or *swipe right* to move between menu items and *double-tap* to select an item. We also included a *single-tap* gesture to repeat a voice prompt, a *swipe-down* gesture to go to the previous menu, and a *tap-and-hold* gesture to select

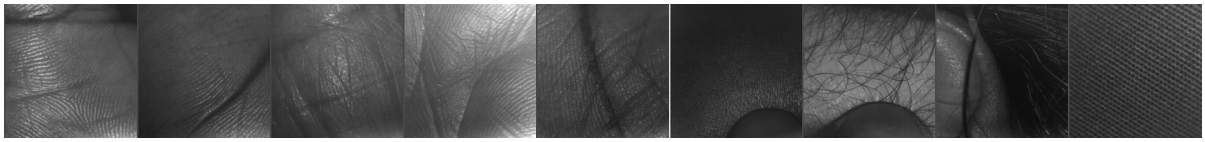


Fig. 7. Sample image data collected with TouchCam Realtime. All images selected from different participants.

[8] http://www.apple.com/accessibility/ios/voiceover/
[9] https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback

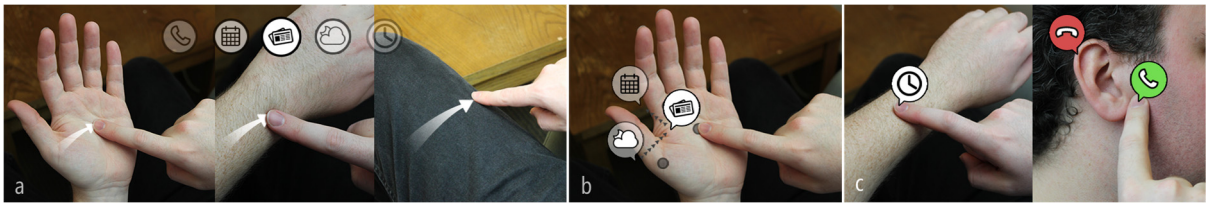
location-specific items. The tap-and-hold gesture was recognized by an 800ms timeout after the *touch-down* event while the other gestures were recognized using a pre-trained SVM classifier (as described in Section 3.2). These gestures can be performed at any body location.

Training Procedure. To limit the amount of time needed to train our system, we implemented an adaptive training procedure inspired by boosting [14]. After capturing a single image of each of the nine locations for initialization, participants then moved their finger around each location in a fixed order as the system continuously classified the video frames. Whenever a video frame was misclassified, that frame and the current location label were saved and the classifiers were retrained. This semi-automated training continued until convergence (*i.e.*, until the researcher determined that the automated system was performing well). After training all locations, at least one additional round was necessary to ensure that new image samples did not negatively affect performance. We found that the initial training images plus two rounds of semi-automated training were sufficient for most users, which took roughly 15-20 minutes and resulted in an average of 13 training examples per location (*SD*=4.5; *range*=5-24).

Procedure. The study procedure lasted up to two hours, and consisted of: (i) an interview about mobile and wearable device usage including thoughts about on-body interaction (~20 minutes); (ii) system calibration and training (~30 minutes); (iii) using TouchCam with three interaction techniques (~10 minutes each); and (iv) a post-study questionnaire (~15 minutes). For (iii), the VoiceOver-like interaction techniques were presented in a fully counterbalanced order. Each interaction technique supported the same set of applications and menu items accessed through a two-level hierarchical menu. The top-level menu had five applications (*Clock*, *Daily Summary*, *Notifications*, *Health and Activities*, and *Voice Input*), which were selected by double tapping. Once selected, each application had 3-4 submenu items except for *Voice Input*, which had no submenu. The three interaction techniques are described below (see also: Figure 8 and the supplementary video):

1. *Location-independent gestures (LI).* Users swiped left or right anywhere to select an application.
2. *Location-specific palm gestures (LS_{palm}).* Top-level applications were mapped to five different locations on the palm. Users pointed directly to a location to select that application or searched for an item by sliding their finger between locations (similar to VoiceOver).
3. *Location-specific body gestures (LS_{body}).* Functioned similarly to LS_{palm} but mapped the applications to five different locations on the body rather than just the palm. We attempted to use intuitive mappings. For example, tapping the outer wrist for Clock and the ear for Voice Input. The other mappings were: the palm for Notifications, the inner wrist for Daily Summary, and the thigh for Health and Activities.

After activating an application, navigation of the submenus was identical across all three interaction techniques, using swipes left and right to select an item and a double-tap to activate it. For each of these interaction techniques, participants were instructed to complete the same set of 10 tasks in a random order. After an automated voice prompt said "begin," a task consisted of selecting an application, opening its submenu, and then selecting and activating a specific menu item (*e.g.*, "open the *Alarm* item under the *Clock* menu"). After



(a) Location-independent gestures (*LI*) (b) Location-specific palm gestures (LS_{palm}) (c)Location-specific body gestures (LS_{body})

Fig. 8. Three on-body interaction techniques: (a) for *LI*, users swipe left/right anywhere on the body to select an application. For (b) and (c), users select an application by double tapping on a specific location on their palm (LS_{palm}) or body (LS_{body}).

the correct menu item had been activated by double-tapping, an automated voice prompt said "task complete" and participants proceeded to the next task.

The session concluded with open-ended questions about the participant's experience using TouchCam Realtime and the three interaction techniques.

Data and Analysis. Throughout the study, we logged all sensor readings, the location and gesture classifications, and event occurrences (*e.g.,* task start/end, menu navigation). We analyze performance in terms of classification accuracy, as well as qualitative metrics of robustness and usability for the three on-body interaction techniques that we tested. We also describe qualitative reactions and subjective preferences based on the interviews and questionnaires.

5.2 Study II: Experiments and Results

To evaluate TouchCam's realtime performance and usability with our blind and low-vision users, we observed participants' behavior during the study and analyzed subjective feedback about our system. We also conducted offline experiments as with Study I, focusing on the sensor data gathered during the training phase of the study (rather than later data, which was unlabeled). Below, we summarize the details of our experiments and findings.

General Observations and Reactions. All twelve participants successfully used TouchCam Realtime to complete tasks with each of the three interaction techniques. In the pre-study interview, most participants (*N*=9) reacted positively toward the idea of on-body interaction citing quick and easy access (N=7), the ability to map specific tasks to different body locations (*N*=6), reducing the number of devices to be carried (*N*=6), and not needing to hold a phone in hand, thus avoiding the risk of theft or damage and potentially freeing that hand for other tasks (*N*=4).

Participants reacted similarly after using the TouchCam prototype. Preferences were split between the three interaction techniques. Participants appreciated the low learning curve and flexible input location of the *LI* interface, which supported simple swipe and tap gestures anywhere on the body, while the location-specific LS_{palm} and LS_{body} interfaces offered quicker and more direct selections once the location mappings were learned. Some participants preferred the proximity of locations for LS_{palm} because it enabled easy exploration and minimal movement, while others liked the more intuitive location mappings of LS_{body}. Key concerns included TouchCam's large physical size, the occasional difficulty with the LS_{palm} interface due to its lower fine-grained accuracy, and the social acceptability of using LS_{body} in public (*e.g.,* touching an ear may draw unwanted attention to the device). See [46] for a more thorough examination of qualitative reactions to our system.

Localization Accuracy. To assess TouchCam's localization accuracy and robustness for visually impaired users, we analyzed the data gathered during the *training* phase of the study. We first conducted a leave-one-out cross-validation experiment using the recorded training samples for each participant (similar to Study I). This resulted in an average accuracy of 91.2% (*SD*=3.5%) at the coarse-grained level and 76.3% (*SD*=76.3%) at the fine-grained level, which is a drop in performance compared to Section 4.3. This decrease, however, is reflective of our adaptive training procedure: since new samples are added only when misclassified using the current SVM, we would naturally expect lower performance when removing even a single sample for cross-validation.

Thus, we conducted an additional experiment using the full training set and classified other video frames from the training session (*i.e.,* those recorded in between the stored training samples). Here, the accuracy increases to 94.2% (*SD*=5.0%) and 81.3% (*SD*=6.6%) respectively. These latter numbers better reflect actual usage performance since we could not reliably measure ground truth during the actual user study (*i.e.,* when participants were using TouchCam with the three interaction techniques). We note that though performance should be improved in future work (see Discussion), these results were sufficient for using and evaluating TouchCam with our participants.

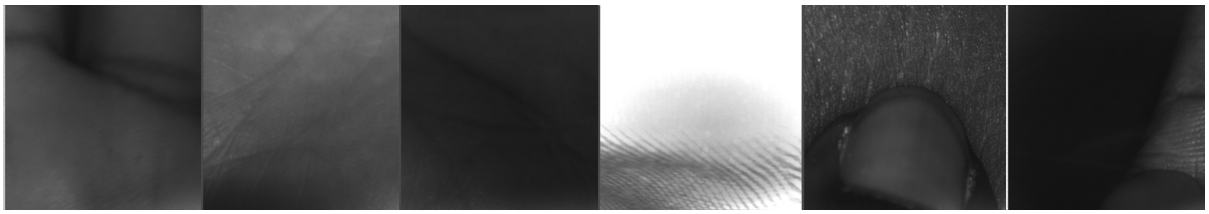


Fig. 9. Some images captured during Study II were of poor quality due to the highlighted reasons. Despite these issues performance remained adequate for participants to complete our specified tasks.

Robustness. To investigate this drop in performance in more detail, we performed a manual inspection of the 1,380 training images across the 12 participants using a custom image reviewing tool. While the severity of the problems varied widely, 22.2% of the images had some issue that could interfere with reliable classification (Figure 9), including: poor focus (13.6%), insufficient illumination (5.4%), poor contrast (4.3%), or oversaturation (0.8%). In addition, 3.2% of the images did not capture the target location due to the offset between the participant's touch location and the center of the camera's field of view, and in 0.6% of the images the participant's finger filled a large portion of the field of view, reducing the number of pixels available for identifying the target location. We further discuss robustness in the Discussion.

5.3 Summary of Study II Findings

Our findings validate TouchCam Realtime's performance with our target population and demonstrate three possible on-body interaction techniques that our approach can support. Participants successfully performed several simple input tasks with our system, and their comments highlight positive reactions to on-body input as well as tradeoffs between the three interaction techniques. These tradeoffs reflect both TouchCam's performance (*e.g.*, LS_{palm} was least accurate due to its reliance on fine-grained localization) and broader design implications (*e.g.*, user preferences for flexibility of input location, learning curve, and social acceptability). Our findings also highlight obstacles to robust on-body input recognition, especially for visually impaired users who cannot rely on visual cues.

6 DISCUSSION

While prior work has explored preliminary issues related to the design of on-body interfaces for visually impaired users [18,44], TouchCam is the first wearable on-body input system supporting real-time interaction designed for and evaluated with this population. Moreover, our work contributes the first real-time system for localizing skin images, and the first to explore location-specific touch-based gestures at a wide set of body locations. Below we discuss TouchCam's performance and usability and provide recommendations for future on-body input systems to support users with visual impairments.

6.1 Robust On-Body Input Detection Using Sensors On the Gesturing Finger and Wrist

Because TouchCam's sensors move with the gesturing finger, they can support touch input at a variety of body (and non-body) locations without requiring additional instrumentation. This feature allows greater input flexibility than most other on-body input approaches (*e.g.*, compared to ViBand [28] or Touché [48]) and means that the user is also less likely to encounter issues with camera framing or occlusion—problems that are common for VI users when they use camera-based systems [1]. Although we did not examine non-body interactions in our work, TouchCam should support location-specific gestures at any surface with visually distinctive features.

Our results demonstrate the feasibility of a computer-vision driven finger camera approach for on-body input; however, we also encountered obstacles that limit TouchCam's accuracy and precision. Because of the

camera's size and positioning, image quality was variable. A high percentage (22.2%) of the training images gathered during Study II were out of focus, low contrast, or poorly illuminated, and in some images the target location was not visible due to the offset between the participant's touch location and the center of the camera's field of view. These usage issues appeared to have a greater impact on performance than other potential factors such as ambient lighting, skin tone, age, or hand size, although future work should investigate these possibilities in greater detail. Improved camera hardware could help address some problems—for example, autofocus functionality would help ensure sharp focus across changes in camera distance or perspective and a wider-angle lens would provide additional contextual information to aid classification. Audio feedback that notifies users when there is a problem and helps them learn how to use the system, as provided by assistive devices for reading such as *KNFB Reader*[10] or *OrCam*[11], could also be helpful. Finally, future work should explore hybrid sensing approaches that combine a finger-mounted camera with an additional body-worn sensor on the head or chest, which could provide additional contextual information and assist with localization.

6.2 An Expanded On-Body Input Vocabulary

As mentioned above, our work introduces new types of on-body interactions that other systems cannot readily support without additional instrumentation. For example, the fixed sensors used by ViBand's smartwatch platform limit interactions to a relatively small area on the hand and arm [28] while Touché requires modification of the target interaction surface and cannot detect gestural input [48]. In contrast, TouchCam can recognize location-specific gestures at several body locations, potentially allowing for intuitive context-specific input (*e.g.*, tapping the wrist to check the time) and supporting a high degree of flexibility and customization.

Participants identified tradeoffs between our three proof-of-concept interface designs, which should be considered when designing on-body interfaces to strike a balance between speed, accuracy, and learnability. Location-independent gestures (*LI*), which allow navigation using swipe gestures anywhere on the body, are easy to understand and learn, do not require individual calibration, and enable flexible input as needed for different situations (*e.g.*, sitting at home *vs.* walking while holding a cane). Location-specific gestures (LS_{palm} and LS_{body}), where the user can directly select an application or menu item by touching a specific location, are potentially quicker once the location mappings have been learned and can also support intuitive context-specific gestures as mentioned above. The palm-only version (LS_{palm}), with its high touch sensitivity and close proximity between mapped locations, could enable faster and more discrete input. Compared with the other two versions, it also more readily supports "touch and explore" functionality that could help participants learn the location mappings more quickly. However, in our experiments LS_{palm} was less accurate than the other two because of inter-class similarity between palm locations and thus required participants to more carefully position their hand and fingers.

This expanded input vocabulary and flexibility of input locations may come at a cost, at least in the current iteration of TouchCam. While our prior work [53] suggested that we should be able to support precise localization on the palm and fingers using their rich visual features, our findings in this work highlight difficulty with robustly recognizing fine-grained locations. Future work should investigate ways to more reliably recognize fine locations, ideally with greater granularity than tested in our studies (*e.g.*, more than five palm locations), and recognizing touch input at two or more locations simultaneously (*e.g.*, using multiple finger-worn sensors) to support multi-touch gestures. In particular, future work should investigate how to extend our approach to support precise 2D localization (*e.g.*, as with OmniTouch [21] or CyclopsRing [4]). These goals may be possible with the aid of additional sensors (*e.g.*, a body-mounted camera) or with more efficient fingerprint and palmprint recognition algorithms that can support real-time interactions.

[10] http://www.knfbreader.com/
[11] http://www.orcam.com/

6.3 Training and Calibration

While TouchCam's gesture recognition algorithms are robust enough to allow for a shared classifier that works across users, its localization algorithms rely on unique skin and clothing features and must be individually calibrated for each user. This requirement raises two concerns: (i) the time needed to complete the individual training procedure, and (ii) the stability and robustness of the classifiers over time as the system shifts position and the user's body appearance varies (*e.g.*, due to changing moisture levels or clothing). We took steps to address the first concern in Study II by introducing our automated training procedure, which took about 15-20 minutes for a new user compared to 30-45 minutes in Study I. However, this procedure will likely need to be simplified and further streamlined in future versions. One possibility would be to bootstrap the system using a large amount training data across multiple users, which could enable coarse-grained classification without individual training. Fine-grained accuracy could be improved over time by learning as the system is used.

As for the second concern, it is possible (even likely) that shifts in the sensor positions after calibration negatively impacted performance for some participants during Study II. Long-term performance is a challenge for many on-body input systems, since they can be highly sensitive to sensor positioning and biometric changes [63]. To explore how accuracy is affected over time, we conducted a small additional study with data gathered across five identical sessions with a single user (the first author). The time between sessions varied from 15 minutes to 24 hours, with the sessions completed over a three-day period. The prototype was fully removed between each session. Classification accuracy was measured similarly to the other experiments described above, except previous session data was used for training and the current session for testing.

As expected, accuracy drops considerably when training on a single session and testing on another, from the 94.2% coarse-grained and 81.3% fine-grained numbers reported in Study II down to 88.2% and 73.6% on average respectively. However, combining training data across sessions improves accuracy reaching an average of 96.5% and 91.8% at the two levels for four training sessions (Figure 10). A larger longitudinal study will be necessary to determine how well these results extend to other users and to a longer period of time, but these results are promising.

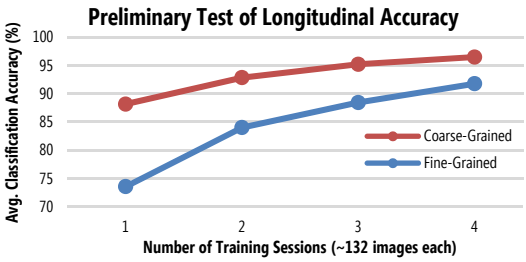


Fig. 10. Classification accuracy across multiple sessions. In general, accuracy increases with more training sessions, suggesting that recalibration may initially be necessary but that accuracy will eventually converge.

6.4 Physical Design

We designed TouchCam to avoid interfering with the user's movements and sense of touch, but the system is still large and requires tethering to a desktop computer for fast processing. With further algorithmic optimizations and increases in mobile processing power, we ultimately envision a smaller, self-contained system that uses a smartwatch for processing and power. Furthermore, our priority with the finger-worn components was to ensure robustness and durability during our experiments, but our design can be streamlined considerably using existing technology. For example, the 6mm diameter camera module[12] that we selected could be replaced with a much smaller 1mm unit from the same manufacturer[13], and the IMU components could be embedded more directly into the ring (while the board we used is 16mm in diameter, the IMU itself is only 4mm square). The IR reflectance sensors positioned near the tip of the user's finger could potentially be replaced with an alternative touch detection method that is less intrusive—for example, an IR depth sensor with a longer range. Further work is needed to explore how these design changes impact accuracy, robustness, and user perceptions.

[12] Awaiba NanEye GS Idule Demo Kit
[13] Awaiba NanEye 2D Sensor

6.5 Limitations

Our system design and studies had several limitations. The TouchCam camera needed to be manually focused and its relatively narrow field of view resulted in an offset between what the user was touching and what was sensed—the latter was particularly problematic for small locations (*e.g.*, finger tips). Future work should explore auto-focusing camera hardware with wide angle lens. The data collected during Studies I and II was collected under controlled conditions. Moreover, while the visually impaired participants in Study II were able to use TouchCam to complete all of the specified tasks, they occasionally needed multiple attempts to do so. Future work should explore more realistic and longitudinal usage.

7 Conclusion

We introduced and investigated TouchCam, a finger-worn, multi-sensor system that supports input at a variety of body locations while mitigating camera framing issues that blind users often experience. Our design also enables new types of contextual gestures based on location. We evaluated two iterations of the TouchCam system in terms of accuracy and robustness, as well as usability for our target group of visually impaired users. Our findings not only highlight the feasibility of our approach—greater than 95% accuracy at detecting 24 location-specific gestures, and support for realtime interaction at approximately 35 frames per second—but also characterize tradeoffs in robustness and usability between different types of on-body input. Fine-grained input on the palm and fingers is desirable for efficient and discrete input, but these locations are more challenging to classify reliably due to their small size and similar visual features; in contrast, disparate body locations are easier to recognize and may enable more intuitive mappings between location and application, but may also be less efficient for a new user and potentially socially unacceptable. Location-specific gestures have the potential to support efficient interaction for expert users, flexible input locations depending on user preference or situation (*e.g.*, while walking with a cane *vs.* sitting at home), task-based interactions tied to intuitive locations, and relatively fine-grained input for body areas that have distinctive visual features (*e.g.*, fingertips and palm). In future work, we plan to explore ways to improve robustness and evaluate our system's long-term performance during a longitudinal study.

8 Acknowledgements

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617. We thank Liang He for his work on the TouchCam video (see supplementary materials) and our participants.

Appendix

Localization Features		**Motion Features**	
IR	2 raw IR sensor readings	IR	70 features: 50 resampled points + 5 summary statistics × 4 windows
IMUs	4D orientation vector (quaternion) for each IMU	IMUs	639 features: 3 sensors × [3 axes × (50 resampled points + 5 summary statistics × 4 windows) + 3 correlation values]
Camera	LBP texture histogram with 1792 bins (14 patterns x 16 variances x 8 scales) 2D Gabor keypoints, variable number per image	Camera	140 features: 2 axes × (50 resampled points + 5 summary statistics × 4 windows)

Table 4. Summary of localization and motion features extracted from each sensor for TouchCam Offline (Study I).

Single Sensors	t_{23}	p	d
IR vs. W	-17.90	<0.001	-3.65
IR vs. F	-17.16	<0.001	-3.50
IR vs. C	-10.38	<0.001	-2.12
W vs. C	4.34	<0.001	0.89
F vs. C	3.63	0.005	0.74

Two Sensors	t_{23}	p	d
IR+C vs. F+W	-8.85	<0.001	-1.81
IR+C vs. W+C	-7.46	<0.001	-1.52
IR+W vs. IR+C	7.39	<0.001	1.51
IR+F vs. IR+C	6.72	<0.001	1.37
IR+C vs. F+C	-6.59	<0.001	-1.35
F+W vs. F+C	6.20	<0.001	1.27
IR+F vs. F+W	-4.67	<0.001	-0.95
F+W vs. W+C	3.73	0.008	0.76
IR+W vs. F+W	-3.23	0.036	-0.66

Best Single (W) vs. Two Sensors	t_{23}	p	d
W vs. IR+C	5.73	<0.001	1.17
W vs. W+C	-4.81	<0.001	-0.98
W vs. IR+W	-4.81	<0.001	-0.98
W vs. F+W	-4.75	<0.001	-0.97

Three Sensors	t_{23}	p	d
IR+F+W vs. IR+F+C	6.46	<0.001	1.32
IR+F+C vs. F+W+C	-4.71	<0.001	-0.96
IR+F+W vs. IR+W+C	3.64	0.003	0.74
IR+F+W vs. F+W+C	-3.17	0.016	-0.65

Best Three (IR+F+W) vs. All Four Sensors	t_{23}	p	d
IR+F+W vs. All	-2.13	0.044	-0.44

Best Two (F+W) vs. Three Sensors	t_{23}	p	d
F+W vs. IR+F+C	4.17	<0.001	0.85
F+W vs. IR+F+W	-2.99	0.020	-0.61
F+W vs. F+W+C	-2.67	0.027	-0.55
F+W vs. IR+W+C	-2.67	0.026	0.49

Table 5. Statistically significant comparisons between sensors used in Study I.

REFERENCES

[1] Dustin Adams, Lourdes Morales, and Sri Kurniawan. 2013. A qualitative study to support a blind photography mobile application. In *Proc. PETRA 2013*: 1–8. https://doi.org/10.1145/2504335.2504360

[2] Daniel Ashbrook, Patrick Baudisch, and Sean White. 2011. Nenya: Subtle and Eyes-free Mobile Input with a Magnetically-tracked Finger Ring. In *Proc. CHI 2011*, 2043–2046. https://doi.org/10.1145/1978942.1979238

[3] John Canny. 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 6: 679–698. https://doi.org/10.1109/TPAMI.1986.4767851

[4] Liwei Chan, Yi-Ling Chen, Chi-Hao Hsieh, Rong-Hao Liang, and Bing-Yu Chen. 2015. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring. In *Proc. UIST 2015*, 549–556. https://doi.org/10.1145/2807442.2807450

[5] Liwei Chan, Chi-Hao Hsieh, Yi-Ling Chen, Shuo Yang, Da-Yuan Huang, Rong-Hao Liang, and Bing-Yu Chen. 2015. Cyclops: Wearable and single-piece full-body gesture input devices. In *Proc. CHI 2015*, 3001–3009. https://doi.org/10.1145/2702123.2702464

[6] Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, Kai-Yin Cheng, Chao-Huai Su, Mike Y Chen, Wen-Huang Cheng, and Bing-Yu Chen. 2013. FingerPad: Private and Subtle Interaction Using Fingertips. In *Proc. UIST 2013*, 255–260. https://doi.org/10.1145/2501988.2502016

[7] Michał Choraś and Rafał Kozik. 2012. Contactless palmprint and knuckle biometrics for mobile devices. *Pattern Anal. Appl.* 15, 1: 73–85. https://doi.org/10.1007/s10044-011-0248-4

[8] Dar-Shyang Lee and S.N. Srihari. 1995. A theory of classifier combination: the neural network approach. In *Proc. ICDAR 1995*, 42–45. https://doi.org/10.1109/ICDAR.1995.598940

[9] Mohammad Omar Derawi, Bian Yang, and Christoph Busch. 2012. Fingerprint Recognition with Embedded Cameras on Mobile Phones. In *Security and Privacy in Mobile Info. and Com. Sys.* Springer, 136–147. https://doi.org/10.1007/978-3-642-30244-2_12

[10] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller, and Max Mühlhäuser. 2012. PalmRC: Imaginary Palm-Based Remote Control for Eyes-free Television Interaction. In *Proc. EuroITV 2012*, 27. Retrieved August 13, 2015 from http://dl.acm.org/citation.cfm?id=2325616.2325623

[11] Murat Ekinci and Murat Aykut. 2008. Palmprint Recognition by Applying Wavelet-Based Kernel PCA. *Computer Science and Technology* 23, 107: 851–861.

[12] Eryun Liu, A. K. Jain, and Jie Tian. 2013. A Coarse to Fine Minutiae-Based Latent Palmprint Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 10: 2307–2322. https://doi.org/10.1109/TPAMI.2013.39

[13] Leah Findlater, Lee Stearns, Ruofei Du, Uran Oh, David Ross, Rama Chellappa, and Jon Froehlich. 2015. Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. In *Proc.* ASSETS 2015, 383–384. https://doi.org/10.1145/2700648.2811381

[14] Yoav Freund and Robert E. Schapire. 1995. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Proc. EuroCOLT 1995*. Springer, Berlin, Heidelberg, 23–37. https://doi.org/10.1007/3-540-59119-2_166

[15] Daniel Goldreich and Ingrid M. Kanics. 2003. Tactile Acuity is Enhanced in Blindness. *J. Neurosci.* 23, 8: 3439–3445. Retrieved November 24, 2013 from http://www.jneurosci.org/content/23/8/3439.abstract

[16] Zhenhua Guo, Lei Zhang, and David Zhang. 2010. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognition* 43, 3: 706–719. https://doi.org/10.1016/j.patcog.2009.08.017

[17] Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. 2010. Imaginary Interfaces: Spatial Interaction with Empty Hands and Without Visual Feedback. In *Proc. UIST 2010*, 3–12. https://doi.org/10.1145/1866029.1866033

[18] Sean G Gustafson, Bernhard Rabe, and Patrick M Baudisch. 2013. Understanding Palm-based Imaginary Interfaces: The Role of Visual and Tactile Cues when Browsing. In *Proc. CHI 2013*, 889–898. https://doi.org/10.1145/2470654.2466114

[19] Sean Gustafson, Christian Holz, and Patrick Baudisch. 2011. Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device. In *Proc. UIST 2011*, 283–292. https://doi.org/10.1145/2047196.2047233

[20] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: Appropriating the Body As an Input Surface. In *Proc. CHI 2010*, 453–462. https://doi.org/10.1145/1753326.1753394

[21] Chris Harrison and Andrew D Wilson. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proc. UIST 2011*, 441–450.

[22] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*: 65–70.

[23] De-Shuang Huang, Wei Jia, and David Zhang. 2008. Palmprint verification based on principal lines. *Pattern Recognition* 41, 4: 1316–1328. https://doi.org/10.1016/j.patcog.2007.08.016

[24] A.K. Jain and Jianjiang Feng. 2009. Latent Palmprint Matching. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 6: 1032–1047. https://doi.org/10.1109/TPAMI.2008.242

[25] Chajndrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. 2011. Supporting blind photography. In *Proc. ASSETS 2011*, 203. https://doi.org/10.1145/2049536.2049573

[26] Wolf Kienzle and Ken Hinckley. 2014. LightRing. In *Proc. UIST 2014*, 157–160. https://doi.org/10.1145/2642918.2647376

[27] Gierad Laput, Robert Xiao, Xiang "Anthony" Chen, Scott E. Hudson, and Chris Harrison. 2014. Skin buttons. In *Proc. UIST 2014*, 389–394. https://doi.org/10.1145/2642918.2647356

[28] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proc. UIST 2016*, 321–333. https://doi.org/10.1145/2984511.2984582

[29] John P. Lewis. 1995. Fast Template Matching. *Vision Interface* 95, 120123: 15–19.

[30] Rong-Hao Liang, Shu-Yang Lin, Chao-Huai Su, Kai-Yin Cheng, Bing-Yu Chen, and De-Nian Yang. 2011. SonarWatch: Appropriating the Forearm as a Slider Bar. In *SIGGRAPH Asia 2011 Emerging Technologies on - SA '11*, 1–1. https://doi.org/10.1145/2073370.2073374

[31] Soo-Chul Lim, Jungsoon Shin, Seung-Chan Kim, and Joonah Park. 2015. Expansion of Smartwatch Touch Interface from Touchscreen to Around Device Interface Using Infrared Line Image Sensors. *Sensors* 15, 7: 16642–16653. https://doi.org/10.3390/s150716642

[32] Shu-Yang Lin, Chao-Huai Su, Kai-Yin Cheng, Rong-Hao Liang, Tzu-Hao Kuo, and Bing-Yu Chen. 2011. Pub - Point Upon Body: Exploring Eyes-free Interaction and Methods on an Arm. In *Proc. UIST 2011*, 481–488. https://doi.org/10.1145/2047196.2047259

[33] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan. 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. In *2011 IEEE International Conference on Rehabilitation Robotics*, 1–7. https://doi.org/10.1109/ICORR.2011.5975346

[34] Heinrich Braun Martin Riedmiller. 1992. RPROP - A Fast Adaptive Learning Algorithm. In *Proc. of ISCIS VII*.

[35] Denys J. C. Matthies, Simon T. Perrault, Bodo Urban, and Shengdong Zhao. 2015. Botential: Localizing On-Body Gestures by Measuring Electrical Signatures on the Human Skin. In *Proc, MobileHCI 2015*, 207–216. https://doi.org/10.1145/2785830.2785859

[36] Abdallah Meraoumia, Salim Chitroub, and Ahmed Bouridane. 2011. Fusion of Finger-Knuckle-Print and Palmprint for an Efficient Multi-Biometric System of Person Recognition. In *2011 IEEE International Conference on Communications (ICC)*, 1–5. https://doi.org/10.1109/icc.2011.5962661

[37] Aythami Morales, Miguel A. Ferrer, and Ajay Kumar. 2010. Improved palmprint authentication using contactless imaging. In *IEEE Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 1–6. https://doi.org/10.1109/BTAS.2010.5634472

[38] Suranga Nanayakkara, Roy Shilkrot, Kian Peen Yeo, and Pattie Maes. 2013. EyeRing: A Finger-worn Input Device for Seamless Interactions with Our Surroundings. In *Proc. AH 2013*, 13–20. https://doi.org/10.1145/2459236.2459240

[39] J Farley Norman and Ashley N Bartholomew. 2011. Blindness enhances tactile acuity and haptic 3-D shape discrimination. *Attention, Perception, & Psychophysics* 73, 7: 2323–2331. https://doi.org/10.3758/s13414-011-0160-4

[40] Masa Ogata and Michita Imai. 2015. SkinWatch. In *Proc. AH 2015*, 21–24. https://doi.org/10.1145/2735711.2735830

[41] Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2013. SenSkin: Adapting Skin as a Soft Interface. In *Proc. UIST 2013*, 539–544. https://doi.org/10.1145/2501988.2502039

[42] Masa Ogata, Yuta Sugiura, Yasutoshi Makino, Masahiko Inami, and Michita Imai. 2014. Augmenting a Wearable Display with Skin Surface as an Expanded Input Area. In *Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments*, Aaron Marcus (ed.). Springer International Publishing, Cham, 606–614. https://doi.org/10.1007/978-3-319-07626-3_57

[43] Masa Ogata, Yuta Sugiura, Hirotaka Osawa, and Michita Imai. 2012. iRing: intelligent ring using infrared reflection. In *Proc. UIST 2012*, 131–136.

[44] Uran Oh and Leah Findlater. 2014. Design of and subjective response to on-body input for people with visual impairments. In *Proc. ASSETS '14*, 8 pages.

[45] Uran Oh and Leah Findlater. 2015. A Performance Comparison of On-Hand versus On-Phone Non-Visual Input by Blind and Sighted Users. *ACM Transactions on Accessible Computing (TACCESS)* 7, 4: 14.

[46] Uran Oh, Lee Stearns, Alisha Pradhan, Jon E. Froehlich, and Leah Findlater. 2017. Investigating Microinteractions for People with Visual Impairments and the Potential Role of On-Body Interaction. In *Proc. ASSETS 2017*, TO APPEAR.

[47] John C. Platt. 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in Large Margin Classifiers* 10, 3: 61–74.

[48] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proc. CHI 2012*, c: 483–492. https://doi.org/10.1145/2207676.2207743

[49] Roy Shilkrot, Jochen Huber, Jürgen Steimle, Suranga Nanayakkara, and Pattie Maes. 2015. Digital Digits: A Comprehensive Survey of Finger Augmentation Devices. *ACM Computing Surveys* 48, 2: 1–29. https://doi.org/10.1145/2828993

[50] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. 2017. WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor. In *Proc. CHI 2017*, 3891–3902. https://doi.org/10.1145/3025453.3026005

[51] Lee Stearns, Ruofei Du, Uran Oh, Catherine Jou, Leah Findlater, David A. Ross, and Jon E. Froehlich. 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *ACM Transactions on Accessible Computing* 9, 1: 1–38. https://doi.org/10.1145/2914793

[52] Lee Stearns, Ruofei Du, Uran Oh, Yumeng Wang, Rama Chellappa, Leah Findlater, and Jon E. Froehlich. 2014. The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind. *Workshop on Assistive Computer Vision and Robotics (ACVR'14) in Conjunction with the European Conference on Computer Vision (ECCV'14)*. https://doi.org/10.1007/978-3-319-16199-0_43

[53] Lee Stearns, Uran Oh, Bridget J. Cheng, Leah Findlater, David Ross, Rama Chellappa, and Jon E. Froehlich. 2016. Localization of skin features on the hand and wrist from small image patches. In *Proc. ICPR 2016*, 1003–1010. https://doi.org/10.1109/ICPR.2016.7899767
[54] Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. 2009. Brainy Hand: an earworn hand gesture interaction device. In *Proc. CHI EA 2009*, 4255. https://doi.org/10.1145/1520340.1520649
[55] Marynel Vázquez and Aaron Steinfeld. 2012. Helping visually impaired users properly aim a camera. In *Proc. ASSETS 2012*, 95. https://doi.org/10.1145/2384916.2384934
[56] Wai Kin Kong and D. Zhang. 2002. Palmprint texture analysis based on low-resolution images for personal authentication. In *Proc. Pattern Recognition '02*, 807–810. https://doi.org/10.1109/ICPR.2002.1048142
[57] Cheng-Yao Wang, Min-Chieh Hsiu, Po-Tsung Chiu, Chiao-Hui Chang, Liwei Chan, Bing-Yu Chen, and Mike Y. Chen. 2015. PalmGesture: Using Palms as Gesture Interfaces for Eyes-free Input. In *Proc. MobileHCI 2015*, 217–226. https://doi.org/10.1145/2785830.2785885
[58] Martin Weigel, Tong Lu, Gilles Bailly, Antti Oulasvirta, Carmel Majidi, and Jürgen Steimle. 2015. iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing. In *Proc. CHI 2015*, 2991–3000. https://doi.org/10.1145/2702123.2702391
[59] Jiahui Wu, Gang Pan, Daqing Zhang, Guande Qi, and Shijian Li. 2009. Gesture Recognition with a 3-D Accelerometer. In *Ubiquitous intelligence and computing (UIC 2009)*, 25–38. https://doi.org/10.1007/978-3-642-02830-4_4
[60] Xiangqian Wu, Qiushi Zhao, and Wei Bu. 2014. A SIFT-based contactless palmprint verification approach using iterative RANSAC and local palmprint descriptors. *Pattern Recognition* 47, 10: 3314–3326. https://doi.org/10.1016/j.patcog.2014.04.008
[61] Xing-Dong Yang, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2012. Magic finger: always-available input through finger instrumentation. In *Proc. UIST 2012*, 147–156. https://doi.org/10.1145/2380116.2380137
[62] Yang Zhang, Junhan Zhou, Gierad Laput, and Chris Harrison. 2016. SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Sking. In *Proc. CHI 2016*, 1491–1503. https://doi.org/10.1145/2858036.2858082
[63] Yang Zhang, Junhan Zhou, Gierad Laput, and Chris Harrison. 2016. SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin. In *Proc. CHI 2016*, 1491–1503. https://doi.org/10.1145/2858036.2858082
[64] Zhi Li, Guizhong Liu, Yang Yang, and Junyong You. 2012. Scale- and Rotation-Invariant Local Binary Pattern Using Scale-Adaptive Texton and Subuniform-Based Circular Shift. *IEEE Transactions on Image Processing* 21, 4: 2130–2140. https://doi.org/10.1109/TIP.2011.2173697

Received May 2017; revised August 2017; accepted October 2017

Design of an Augmented Reality Magnification Aid for Low Vision Users

Lee Stearns
University of Maryland
College Park, MD, USA
lstearns@umd.edu

Leah Findlater
University of Washington
Seattle, WA, USA
leahkf@uw.edu

Jon E. Froehlich
University of Washington
Seattle, WA, USA
jonf@cs.washington.edu

ABSTRACT

Augmented reality (AR) systems that enhance visual capabilities could make text and other fine details more accessible for low vision users, improving independence and quality of life. Prior work has begun to investigate the potential of assistive AR, but recent advancements enable new AR visualizations and interactions not yet explored in the context of assistive technology. In this paper, we follow an iterative design process with feedback and suggestions from seven visually impaired participants, designing and testing AR magnification ideas using the Microsoft HoloLens. Participants identified several advantages to the concept of head-worn magnification (*e.g.,* portability, privacy, ready availability), and to our AR designs in particular (*e.g.,* a more natural reading experience and the ability to multitask). We discuss the strengths and weaknesses of this AR magnification approach and summarize lessons learned throughout the process.

Author Keywords

Low vision; visual impairment; assistive technology; wearables; augmented reality; magnification.

ACM Classification Keywords

• Human-centered computing → Accessibility technologies
• Computing methodologies → Mixed / augmented reality

INTRODUCTION

Recent advancements in augmented reality (AR) have the potential to increase the quality of life for people with visual impairments (VI). For low vision users, head-mounted displays (HMDs) that enhance a user's existing visual capabilities are particularly promising. For example, *ForeSee* [25] used an *Oculus Rift* VR headset with an attached camera to magnify and enhance text content, and other researchers have used *Google Glass* to enhance edges within the wearer's field of view [9] or display magnified content from a smartphone screen [16]. Several commercial HMDs (*e.g.*, *eSight* [27], *NuEyes* [28], *IrisVision* [29]) display magnified video captured from a head-mounted camera, and provide image enhancement features such as contrast adjustment. A recent study investigating the use of one of these systems (eSight) was generally positive, showing the impact HMDs can make in users' lives [26].

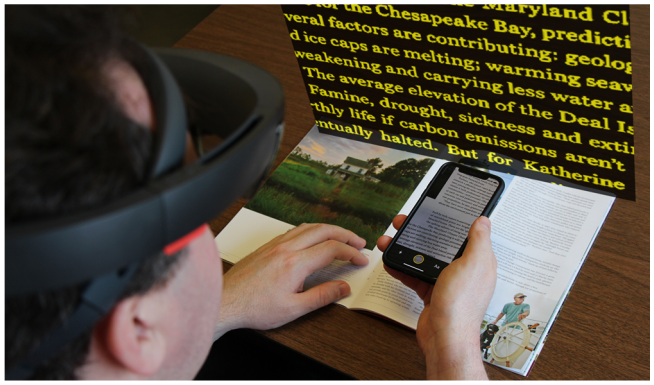


Figure 1. Prototype AR Magnification system using a transparent HMD (the Microsoft HoloLens) and a handheld smartphone (iPhone X) as a camera and input device.

While these systems have begun to explore how HMDs and wearable cameras can be used to augment visual perception, they are limited to enhancing and/or magnifying the 2D image from a video camera. In contrast, the classical definition of AR integrates 3D virtual objects into the 3D physical environment [1], which would allow for new visual enhancement possibilities that are better integrated with the user's real-world tasks. For example, a magnified view of an object can be rendered directly on top of the real object, fixed to a desk near the user's primary work focus, or "projected" on a nearby wall. Off-the-shelf technologies such as the *Microsoft HoloLens* [30], an optical see-through display, are beginning to have the capability to support these types of 3D AR designs.

To investigate the design possibilities for AR magnification tools enabled by registering virtual content in real 3D space, we conducted a series of design sessions with seven low-vision participants. We developed initial prototype designs on a Microsoft HoloLens, which we presented to participants to solicit feedback and open-ended ideas about future wearable magnification aids. Our designs explored several different virtual display options (*e.g.,* affixed to real objects *vs.* moving with pointing finger), image acquisition

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ASSETS '18, October 22–24, 2018, Galway, Ireland
© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-5650-3/18/10 $15.00
https://doi.org/10.1145/3234695.3236361

methods (head-mounted, finger-mounted, or smartphone), and interaction techniques (*e.g.,* Figure 1). The designs were updated between sessions based on participant feedback as well as our own observations.

Overall, participants liked the concept of a wearable AR magnification aid, especially the natural reading experience and ability to multitask that the projected 3D renderings enabled. At the same time, our system presented some difficulties compared to participants' existing magnification aids. We discuss these issues along with potential solutions and design implications.

Our contributions include: (i) an exploration of the design space for augmented reality magnification; (ii) proof-of-concept implementations evaluated and refined through iterative design with low vision users; and (iii) common themes and recommendations that should inform the design of future AR vision enhancement aids for low vision users.

RELATED WORK

Our research is informed by prior work on magnifiers, AR systems, and wearables designed for low vision users.

Low Vision Magnification Aids

Common magnification aids for low vision users include low-tech optical magnifiers, desktop video systems (*e.g.,* [31,32]), handheld devices (*e.g.,* [33,34]), and more recently mobile applications that use a smartphone's camera and display (*e.g.,* [35]). Both Virgili *et al.* [23] and the *American Foundation for the Blind* [2,3] provide a comprehensive overview of these options.

Desktop video magnifiers, also known as closed-circuit television (CCTV) systems, project images from a stationary or handheld camera onto a large television or computer monitor. Some examples are the *ClearView+ HD* [31] and the *DaVinci HD* [32]. These systems benefit from a large screen and ready customization, allowing users to adjust the positioning, magnification level, and other features for a comfortable reading experience. However, they are not portable and the magnified display is separated from the actual content, which could cause difficulty when reading longer passages or searching for a specific location at higher magnification levels. To address this latter issue, a conventional optical magnifier or other digital magnification aid is sometimes used to locate a target before reading it with the CCTV [4].

Handheld video magnifiers are a portable alternative that can be used in a greater variety of situations (*e.g.,* to identify items on a shelf). For example, *Ruby* [33] and *Compact+ HD* [34] both function as a magnifying glass with image enhancement capabilities. Dedicated hardware can be optimized for the needs of low vision users (*e.g.*, large, high-contrast components and controls) and physical buttons allow for adjustment of common settings (*e.g.,* brightness and contrast). However, most of these systems are designed for a single purpose, so users may need to carry multiple devices to identify both near and distant text or other objects, in addition to other general devices such as smartphones. Also, despite theoretically being portable, many of these devices are still bulky, necessarily occlude the physical object being magnified, and may need to be held in such a way that viewing the magnified image cannot be done discreetly.

Smartphones also increasingly support magnification for low vision users. Apple and Android phones now include a built-in magnifier [35,36] and numerous free or low-cost third-party apps are available in the Apple or Google stores. Unlike dedicated handheld magnifiers, these smartphone-based solutions are cheaper, do not require carrying an additional device, and are likely to draw less attention to the user. However, they present other similar disadvantages to dedicated handheld magnifiers.

Our research builds on these designs, using augmented reality to provide a portable and perceptually large virtual display with common image enhancements (*e.g.,* brightness, contrast, color alteration) that can be used to identify text and other fine details, both near and far.

HMDs to Enhance Vision Capabilities

Head-worn vision enhancement systems for low vision users were first proposed in the 1990s [7,12]. Compared to other types of low vision aids, HMD-based solutions offer potential advantages of portability, ready availability, and privacy while displaying enhanced information within the wearer's field of view. A recent study by Zolyomi *et al.* [26] showed that one such device (eSight [27]) improved access to information and social engagement but also had negative social impacts [26]. Another study by Profita *et al.* [15] investigated the social acceptability of HMDs, showing greater acceptance if the device is being used for an assistive purpose as opposed for a general mobile computing task.

Most HMD-based systems for low vision users project magnified and/or enhanced 2D video captured from a wearable camera onto screens mounted in front of the user's eyes [12,25,27–29]. Some recent examples have used consumer VR hardware: *ForeSee* [25] uses an Oculus Rift headset and *IrisVision* [29] uses a head-mounted smartphone (Samsung GearVR). Optical see-through displays have also been employed for vision enhancement [9,16], where virtual information is overlaid on a transparent display, thus augmenting rather than replacing the user's vision. For example, Google Glass has been used to display a magnified view of a smartphone screen [16] and to overlay enhanced edges onto the wearer's view of the real world [9]; however, Glass itself is a low-resolution display (640×360) and not designed as an AR device (*e.g.,* the display is positioned in the user's visual periphery).

None of the above solutions have overlaid virtual content in 3D space—our focus. However, Zhao *et al.* [24] conducted an accessibility evaluation of the *Epson Moverio BT-200* smart glasses with participants with low vision. They

concluded that while the semi-transparency of optical see-through displays did reduce contrast and make it somewhat harder for low vision users to read text or identify shapes, participants were able to successfully use the device and were positive about the experience, confirming that devices like the HoloLens are a useful prototyping platform for our investigation. In a recent ASSETS poster [20], we presented three early designs exploring how the HoloLens might support low vision users by enhancing an image captured by a finger-mounted camera: showing a magnified view of the camera image at a fixed position in the wearer's view, affixing the image to a 3D surface, and having the image track the user's finger. While illustrating some possibilities for true AR vision enhancement, that work was preliminary did not include any evaluation or feedback from low vision users. Our current research builds on these early explorations, exploring the capabilities of the HoloLens as an AR magnification aid and evaluating several designs with low vision users.

Other Assistive Wearables for Low Vision Users

Our work also builds on other research on assistive wearables for blind and low vision users. The majority of these tools are designed for users with a wider range of visual abilities, including no functional vision, and so do not include a display. Information is primarily conveyed using text-to-speech and other audio cues.

OrCam [14] is a commercial head-worn camera system (with no visual display) that users control with speech and gestures. The system can identify and read printed text using text-to-speech and can identify pre-trained products and faces. Other systems use cameras worn on the head [11] or wrist [19] to identify faces, and an early version of Microsoft's recent *Seeing AI* smartphone app [37] used a pair of smart-glasses to identify people, describe facial attributes, and read text, and was controlled by touch gestures on the side of the glasses [38]. Finger-mounted cameras have also been used to support visually impaired users in a variety of tasks, including reading printed text via text-to-speech output [18,21], controlling mobile devices [22], and identifying currency or other objects [13].

While our work differs from these systems in application and in some cases target user group, we also explore tradeoffs between wearable/portable camera locations, including head-worn, finger-worn, and handheld.

A DESIGN SPACE FOR MAGNIFICATION AIDS

To inform the design of an AR magnification aid for low vision users, we first outline our goals and important design dimensions for mobile and/or wearable magnification aids.

Design Goals

Informed by prior work, existing commercial systems, and our own experience working with initial AR prototypes, we formulated the following design goals for our study:

- **Augment rather than replace.** Whenever possible, avoid interfering with the user's existing vision capabilities. Provide enhanced content alongside the real world and provide easy controls to hide or reposition the digital information as needed.
- **Leverage augmented reality.** Go beyond the static 2D displays provided by existing systems and explore applications for persistent digital content overlaid in 3D onto the physical world.
- **Prioritize customization and flexibility.** To support a wide range of vision levels and different situations, the ability to customize how the enhanced content functions and appears is crucial [10].

Design Dimensions

To achieve these goals, we considered several design dimensions in addition to virtual display position, our primary dimension of interest:

- **Virtual display position.** The ability to anchor virtual content to a physical location in 3D space enables several possible virtual display designs. Specifically, we explore four positions. The first, simplest position is a fixed heads-up display that moves with the user's head to always stay within their field of vision. The second position is a stationary display attached to a location in the physical world, which maintains its position as the user moves. The third option is a dynamic display that acts as a magnifying glass and follows the user's hand or other moveable object (*e.g.,* a ring or smartphone). Finally, the fourth position projects an image directly onto the physical object that is being enhanced (*e.g.,* a magnified view shown atop a document).
- **Content capture.** To capture video for processing and display, possible camera locations include head-worn (*e.g.,* [14,25,27–29]), hand-held (*e.g.,* [33,34,37]), and finger or wrist-worn (*e.g.,* ring or smartwatch; [18–20]). We explore a few of these options and discuss the advantages and disadvantages of each.
- **Image enhancements.** To support a range of vision levels, important enhancements include magnification, changes to brightness and contrast, binary thresholding, and color alterations (*e.g.,* as described in [25]). Although not the focus of our study, optical character recognition could also be useful, either to read text aloud or to visually enhance detected text by increasing the resolution or replacing fonts.
- **Physical HMD.** Several display types have been explored previously, as discussed in the Related Work section. However, an optical see-through display and 3D sensing capabilities are needed to achieve our design goals, making the Microsoft HoloLens the obvious choice at the time this research was conducted. The HoloLens allowed us to rapidly prototype and iterate on AR designs; however, we fully expect that future HMDs for AR will be more streamlined, lightweight, and portable (*e.g.,* integrated into traditional glasses).

- **User input.** To support our goal of customizability and flexibility, the AR system needs to provide intuitive and easy-to-use interactive controls. A few options include physical controls on the device or a separate remote (*e.g.,* eSight [27], Glass [5]), gaze tracking, midair gestures, and voice commands (*e.g.,* OrCam [14], HoloLens [30]), or 3D tracking of a physical object (*e.g.,* Oculus Rift controller [39]). We explore a few of these options to see how well they work for low vision users and in different situations.

ITERATIVE DESIGN OF A PROTOTYPE SYSTEM

To explore these design dimensions and evaluate which designs and features would work best for low vision users, we conducted a series of iterative design sessions. These sessions were structured to elicit general feedback and open-ended design ideas from participants, drawing on elements of cooperative [6] and participatory [17] design methodologies. Based on ideas from existing magnification aids, knowledge of available hardware capabilities, and our own experience working with low vision users, we developed an initial prototype system that implemented several user interface designs. We then asked participants to use the system and provide feedback, refining our design over nine design sessions with seven unique participants (two participants returned for a second session). While we modified the system between sessions to fix issues and make minor improvements, for ease of presentation, we group our prototypes into three basic designs based on the broad design elements, components, and user interactions.

Initial Investigation: HoloLens Only

Our initial design used only the HoloLens headset. As mentioned earlier, the HoloLens includes an optical see-through display on which translucent virtual objects ("holograms") can be overlaid in real 3D space. The estimated field of view is 30° × 17.5° with 2500 light points per radian. Microsoft's APIs include motion tracking and 3D scene analysis functions that allow developers to anchor digital content to a physical location in the world so that it will remain stationary as the user moves. The APIs also support input using midair gestures, the direction the user's head is pointing, and voice.

This initial prototype used the HoloLens's built-in camera to capture images of what the user looked at and provided two modes for displaying an enhanced version of those images: fixed 2D and fixed 3D. While we describe these display modes in more detail in the next section, the fixed 2D display moved with the user's head to always remain within view while the fixed 3D mode was anchored to a surface in the physical world. Users could toggle between modes using voice commands and two image enhancement options were provided: binary thresholding (black text on a white background) and color inversion.

While this initial design was functional, internal testing revealed that the HoloLens' built-in camera resolution was simply too low to be useful when magnified. Additionally, requiring users to turn their head to look at desired content for magnification was uncomfortable, and the voice commands were cumbersome and imprecise. We used these observations to develop the next iteration of our prototype, which was the first to be tested with low vision participants.

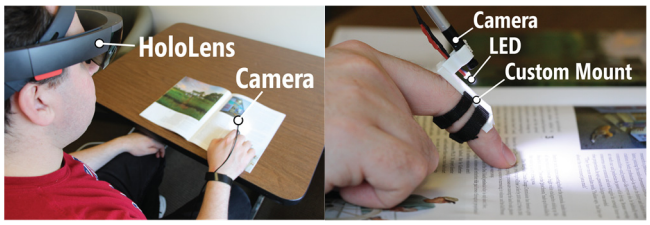


Figure 2. First prototype system design: (a) full system with the HoloLens, (b) close-up of the finger-worn camera

Prototype 1: HoloLens and Finger-Worn Camera

To address the issues observed in our initial investigations and to expand on our design, we added an external camera, implemented two additional display modes and more customization options, and replaced the voice commands with a virtual menu controlled using midair gestures. We then conducted design sessions with three participants, making minor changes to the system between sessions based on feedback (*e.g.,* modifying the perceptual distance at which the AR displays were drawn, simplifying and polishing user input).

Implementation Details

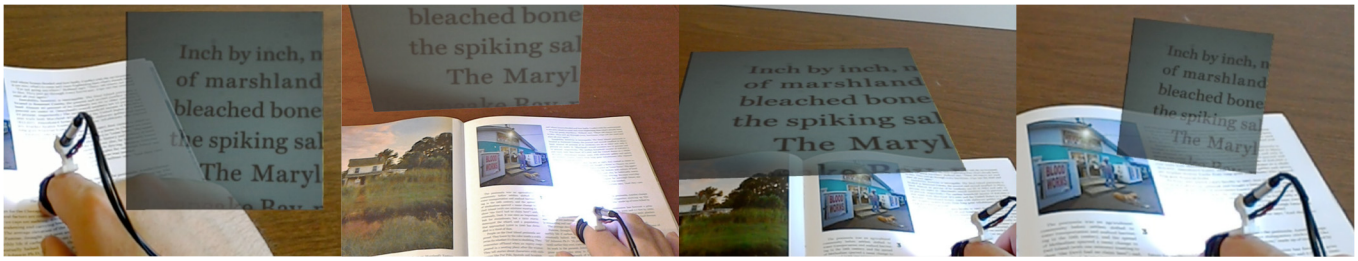
Below we describe the prototype's components and physical design, the four display modes, and the user interactions.

Hardware and Physical Design. This prototype used an external camera[1] mounted on the user' finger using a custom 3D-printed ring with Velcro straps and an LED to provide consistent lighting (Figure 2). The camera provided a close-up view of the target content and allowed the user to read without needing to frame the text within the head-worn camera's field of view. As discussed earlier, similar wearable cameras have been used for other assistive devices [13,18,21,22], albeit primarily for people with more severe visual impairments. We used a laptop computer to capture and process images from the camera, which we streamed wirelessly in real-time to the HoloLens for display.

Virtual Display. To elicit feedback on a range of AR display designs, we implemented four display options for the enhanced view from the camera, including the two explored in the initial prototype (Figure 3 and video figure):

- *Fixed 2D:* This design displayed the image at a fixed location relative to the user's head, which meant that the display was within the user's view at all times. This design is similar to past work using HMDs for visual enhancement (*e.g.,* [9,25,27]), and it is possible to implement on more basic HMDs such as Google Glass.

[1] Awaiba NanEye GS Idule Module, 640×640px, 30° FoV



(a) Fixed 2D (b) Fixed 3D (Vertical) (c) Fixed 3D (Horizontal) (d) Finger Tracked

Figure 3. Prototype 1 provided four virtual display modes, which could be customized (position, size, zoom) using midair gestures. See the accompanying video figure for a demonstration.

- *Fixed 3D Vertical:* This design allowed the user to place the enhanced view from the video camera at a fixed position in the physical world. The display was oriented vertically, allowing for placement on a vertical surface like a wall. It remained at the fixed location as the user turned their head or moved around.
- *Fixed 3D Horizontal:* This design was similar to Fixed 3D Vertical but the display was oriented horizontally so it could be placed on a tabletop or other flat surface.
- *Finger Tracked:* In this design, the display was oriented vertically above the user's finger and moved along with the user's hand like a magnifying glass. Because the HoloLens APIs could not detect the user's hand when touching a page, we used a simple image processing technique to detect the bright LED from the finger-mounted camera and position the display near it.

As with the previous prototype, the system provided image binarization and color inversion features, which participants could use as desired. The brightness of the LED and the HoloLens display could also be adjusted.

User Interactions. Because voice commands proved too limiting for the variety of customization options we wanted to support, we used the gesture recognition capabilities provided by the HoloLens to allow users to adjust the display's position, size, and zoom level. Users opened the virtual menu using the default HoloLens "air tap" gesture, and selected from three options (position, size, or zoom level) by turning their head to position a cursor atop the desired item and then performing another air tap gesture. To move or adjust the display, users performed a "manipulation" gesture, first lowering their index finger, moving their hand in any direction, then raising their finger again once satisfied.

The first two VI participants found these interactions to be difficult, so we reduced the number of menu items (leaving only position and size) and added remote controls to allow us adjust settings as directed by the participant if needed.

Design Sessions

Three participants used our first prototype, comparing the display modes and discussing the overall idea of augmented reality vision enhancement.

Participants. We recruited three participants who used a CCTV or other magnification aid in their daily lives (two male, one female, age range 28–54). The cause and severity of participants' visual impairments varied (Table 1).

Methods. After an open-ended interview to collect demographic information and participants' experience with magnification aids, we introduced our system and demonstrated its features. Participants then used each of the four display modes in a partially counterbalanced order (using 3 out of 4 orders from a balanced Latin square) to read text on a variety of objects, including simple printed documents as well as mail, a pill bottle, and a box of cereal. After each mode, participants provided feedback on their likes and dislikes for that particular mode, as well as thoughts about the customization options. At the end of the study, we asked about experiences using the system and which display modes were most and least preferred, discussed the overall design of the system, and elicited

ID	S1	S2	Age	Gender	Diagnosis	Visual Acuity (self-reported)	Visual Field
P1	✓	✓	28	M	LHON	20/400 or 20/450	Limited central vision
P2	✓		46	F	Retinitis pigmentatosa	Low vision (acuity unknown)	Limited
P3	✓	✓	54	M	Optic atrophy (meningitis)	20/200	Full
P4		✓	29	F	Tumor	Low vision (acuity unknown)	None in left, tunnel vision in right
P5		✓	58	M	Cone-rod dystrophy	Light and shapes (acuity unknown)	Limited central vision
P6		✓	33	F	Oculocutaneous albinism	20/400 in good lighting	Full
P7		✓	68	F	High myopia	20/400, better in ideal conditions	Full, but better peripheral vision

Table 1. Demographic information for the participants across all design session. Columns "S1" and "S2" indicate participation in design sessions with prototype 1 and prototype 2, respectively.

suggestions for improvements and new features. Each session lasted approximately two hours, and participants were compensated $60 for time and travel costs.

Overall Response and Display Modes. The participants each used the system to read the provided materials, with varying levels of success. P1 and P3 reacted positively to the concept of a wearable AR magnifier. P1 commented:

> *"If there was something I could just wear and then be able to see something better, point the camera at it and then have it right there in front of my eyeball then I would use that all the time… You could certainly do many things that you're not able to do by yourself at this point."*

Both P1 and P3 observed advantages to the 3D design elements incorporated into three of the display modes. They considered the two fixed 3D display modes to be more like the reading experience with a CCTV or handheld magnifier than the other two designs, while the dynamic finger tracking design could potentially help to quickly locate a particular location in a document.

Overall, P1 preferred the two fixed 3D designs (vertical or flat) because they were steadiest and easiest to read, while P3 preferred the fixed 2D design because it was always visible and required the least concentration to use. In contrast to the other two participants, P2 found the reading process too difficult and did not see advantages to the AR magnification approach, stating that she would prefer to use audio output from a screenreader for most reading tasks.

All three participants disliked the dynamic finger tracking display, primarily due to technical issues with our implementation. This design required participants to turn their head to look directly at their finger, which had two problems: first it was uncomfortable and required additional concentration to keep their finger always within the HoloLens camera's field of view, which interfered with the reading experience. Second, the bright LED reduced the contrast of the virtual display and made it more difficult to read the enhanced text. Interestingly, P1 also found the Fixed 2D display to be difficult to use because its perceptual distance was fixed too far away—we made this setting adjustable for future participants.

Finger-worn Camera. Perceptions of the finger-worn camera were also mixed. The camera was wearable, which allowed for hands-free use, and separation from the display provided flexibility to allow participants to find a comfortable reading position. However, participants disliked the need to move their finger to read (P2) or found it difficult to move from one line to the next for longer passages (P3). The biggest limitation was the small field of view due to the camera's proximity to the page—each image contained only 3–4 lines of text and a few words across. This problem was compounded by the limited field of view of the HoloLens, which when magnified to a readable size sometimes meant that participants could only fit a word or two on the display at a time. All three participants mentioned that their existing magnification aids had a similar problem, but also stated that our system was worse in its current implementation.

The HoloLens Device. Participants' other feedback about the prototype system primarily centered around limitations of our chosen hardware, especially the physical size, weight, and display contrast. Contrast was a source of difficulty for all three participants, to varying degrees. Images displayed on the HoloLens screen are translucent, which meant that text could be difficult to recognize depending on the background imagery. This issue was addressed somewhat by lowering the room lighting or moving the display so that it was positioned over a flat, high contrast surface (*e.g.,* a white wall or black screen). As mentioned above, the bright LED interfered with reading, so the participants mostly positioned the virtual displays so that they were not looking directly toward it. Even with these measures, P2 was unable to successfully use the system to read because of how the HoloLens display functioned, only able to make out a few scattered words and letters. This finding fits with previous mixed results using optical see-through displays for low vision users [24], and suggests that the HoloLens may work better for some types of visual impairment than others.

User Input. While P1 and P3 were able to use the midair "air tap" gestures to adjust the display, all three participants found the gestures to be cumbersome and difficult to use. We frequently needed to provide assistance when changing settings. Because of these difficulties, participants may not have fully customized the system to meet their specific needs. Additionally, the combination of the slow input gestures and the camera's physical positioning meant that participants could not quickly adjust the magnification level to help with locating the start of a new line or another desired location in a document.

Technology Comparisons. When asked to compare the device with their existing magnification aids, all participants stated that the current version was less convenient, primarily due to limitations with the physical hardware. However, if those issues could be solved, P1 stated:

> *"In comparison to the portable CCTV I have or the full size one, this would be something you could wear and take with you… If you just have a pair of glasses that could essentially do the same thing [as a phone] then I would probably use that even more than my phone."*

Summary. Two participants reacted positively to the idea of AR magnification and observed potential advantages to our 3D display modes. Hardware and user interface issues—especially the field of view, image contrast, and midair gestures—limited the usability of our prototype, with one participant unable to use the system to read at all. Despite these issues, the overall concept showed promise.

Prototype 2: HoloLens and Smartphone

To address these Session 1 study findings, we redesigned several aspects of our system (detailed below). We then recruited six participants for further design sessions using the updated prototype.

Implementation Details

Below we describe changes to the prototype's components and physical design, display modes, and user interactions. Figure 4 shows the updated prototype in action.

Hardware and Physical Design. Because the finger-worn camera had a narrow field of view and required manual focus, we decided to instead experiment with a handheld smartphone camera (an iPhone X). The smartphone is heavier than the finger-worn camera and does not allow for hands-free usage—a feature of the previous design that participants found appealing—but the change provided a higher-quality camera with built-in processing and wireless communication capabilities. In particular, the camera's high resolution and autofocus allowed users to easily control the amount of text captured by moving the phone toward or away from the page. Users could also adjust the brightness of the phone's camera flash LED to help with contrast. We imagine that a future wearable device (*e.g.,* ring or smartwatch) could incorporate these features as well, if they proved useful for applications like this one.

The use of a smartphone also enabled several new user interactions to control the display settings using the touchscreen and motion sensors, which we discuss below. The phone connected wirelessly with the HoloLens to stream video, touchscreen gestures, and 3D motion data.

Virtual Display. Our updated prototype provided three display modes (Figure 5; video figure), which were refined versions of the four tested previously. We differentiated the modes by the object to which the display was attached:

- *Attached to Headset.* The display mode was based on the Fixed 2D mode described previously, but with finer-grained control over the relative position and angle of the display. Users could place the display at a location in front of them, and it would move and rotate with them as they turned their head or moved their body, always maintaining the same relative position.

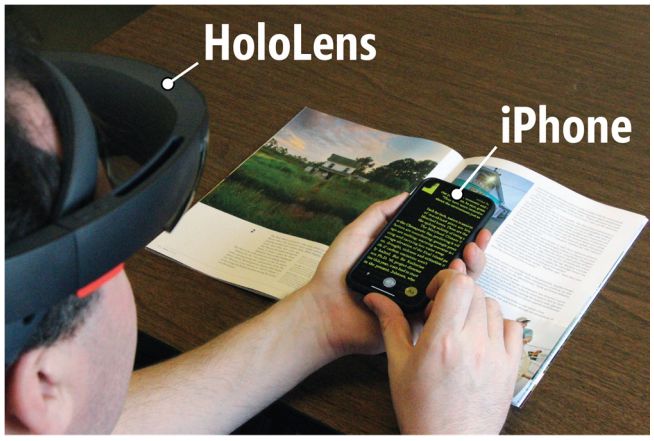
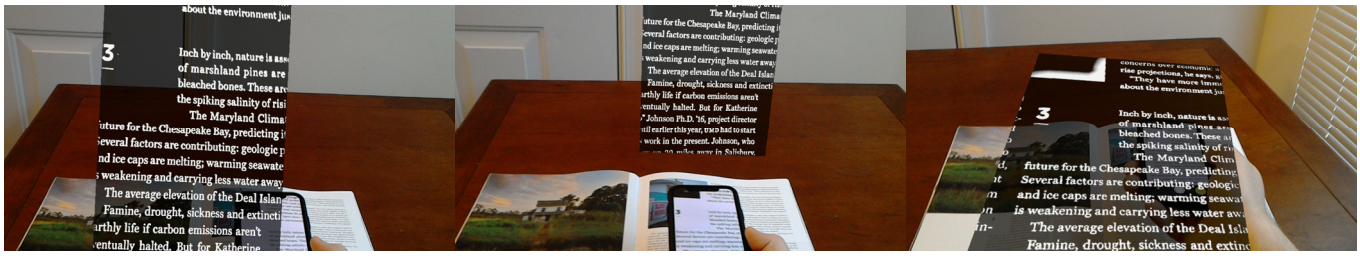


Figure 4. Second prototype AR magnification system using the HoloLens and a hand-held iPhone X

- *Attached to World.* This mode combined the two Fixed 3D designs from the previous study into a single flexible approach that allowed users to position the display in the physical world at any location and angle. As with the earlier designs, the display remained fixed at that physical position as the user moved.
- *Attached to Phone.* This mode functioned similarly to the Finger Tracked design from the previous prototype. It positioned the display atop the smartphone and the display moved as the user moved the phone, acting like a magnifying glass but with an arbitrarily large virtual display. Users could reposition the display—for example, so that it would be vertical while they held the phone horizontally—but the display would always move to maintain the specified position and rotation relative to the phone.

Aside from the three display designs, we also provided controls for users to adjust the image colors and contrast. As before, we made changes to our designs between sessions. For example, in addition to the black and white color inversion options included for the previous prototype, at the request of P4 we implemented standard white/blue, yellow/blue, grayscale, yellow/black, and red/black filters provided by other digital magnifiers. To address comments from P1 and P3, who were the first to try the new prototype, we also added a "freeze frame" mode which allowed users to press a button to stop the video capture and send a full-resolution photograph to the HoloLens for display. Users



(a) Attached to Headset **(b) Attached to World** **(c) Attached to Phone**

Figure 5. Prototype 2 provided three virtual display modes, which were refined versions of the four included with Prototype 1. See the accompanying video figure for a demonstration.



Figure 6. Touchscreen controls on the iPhone prototype. Left to right: main screen, display mode menu, text colors menu.

could then control the image size and position on the AR display as before, but without needing to hold their phone above the target content while reading.

User Input. Touchscreen controls were used for most input (Figure 6), including: double-tapping to open the display mode menu, tapping to select buttons on the screen, pinching to control the size of the virtual display, and sliding to move the display during "freeze frame".

To control the display's 3D position and rotation, we implemented a motion tracking feature using the iPhone's built-in *ARKit* APIs [40]. The API provides functions to track the phone's 3D pose relative to its starting location, which we stream to the HoloLens and use to position and rotate the virtual display. Because the iPhone and HoloLens had different internal 3D reference frames, we manually initialized the transformation between the two at the start of each session using a simple procedure that required visually positioning the phone atop a virtual representation. This procedure is overly simplistic, and a more robust method will likely be necessary for long-term use. However, it proved sufficiently reliable for the duration of our study.

Users could move the phone to position the virtual screen as desired for each of the three display modes. In both the Attached to Headset and Attached to World modes, users pressed a finger firmly on the screen until the phone vibrated, moved the virtual screen to the desired location (3D position and rotation), then released their finger. The interaction was slightly different for the Attached to Phone mode, with users first moving the phone to where they wished the screen to be located, then pressing firmly and moving the phone to where they wished to hold it while reading. After lifting their finger off the touchscreen, the virtual screen maintained the offset between the initial and final positions as they continued to move the phone.

Design Sessions

Six participants tested our updated prototype, comparing the display modes and providing general feedback and open-ended suggestions about their ideal magnification aid.

Participants. We recruited six people with visual impairments (3 male, 3 female, ages 28–68) to participate in design sessions with our updated prototype. P1 and P3 returned from the previous sessions, while four participants had not used our prior prototypes. As with the previous design sessions, the cause and severity of the participants' visual impairments varied (Table 1) but all participants regularly used some type of magnification aid.

Methods. The user sessions were structured similarly to the previous ones. Participants were introduced to the updated prototype and allowed time to explore the options while reading a simple document. After becoming comfortable with the controls, participants then used each of the three display modes in a fully counterbalanced order to read text on a variety of objects, including simple printed documents, magazine articles, mail, and product labels (*e.g.,* box of cereal, pill bottle). After each mode, participants provided feedback on what they liked and disliked. The session closed with a discussion of participants' overall experience using the system, preferred display modes, and feedback on the system and AR magnification in general, and participants' envisioned ideal magnification aid. As with the previous stage, sessions lasted approximately two hours, and participants were compensated $60 for time and travel.

Overall Response. Participants were in general more successful and positive about the experience of using this prototype than we had observed with the previous version. The iPhone provided higher quality images and better control over the amount of visible text, and the touchscreen and motion controls provided faster and easier control over enhancements and virtual display settings. Participants were better able to experience the augmented reality aspects of our approach, which most participants found promising. One participant was particularly enthusiastic about the Attached to World design, stating that it was:

> *"so much better [than her CCTV], you can go down the whole page and read it. Like if I want to read a book or something to my kids, Mommy doesn't have to go line by line. I can read it and keep the flow going. You can read your whole mail in its entirety without it being on your TV."* (P4)

She felt that our system provided an experience more like what she remembered before her vision loss with advantages of portability and privacy compared to her existing aids, continuing, "*It's everything I need as far as being able to read independently"* (P4).

Virtual Display Modes. Participants' display preferences were again mixed, with some participants stating that they liked the flexibility of having multiple designs available and would use different versions depending on the situation. P1 and P3 preferred the Attached to Headset design because they found it easier to focus on the text with fewer variables to consider. In contrast, P4 found that mode to be too distracting, especially when speaking with someone or otherwise multitasking, and preferred the Attached to World design since it functioned, *"like a private, portable CCTV that stays where you want it to stay"* (P4). P5, P6,

and P7 saw advantages to all three designs, including the simplicity of the Attached to Headset design, the natural reading experience and ability to multitask with the Attached to World design, and the versatility and intuitive interactions of the Attached to Phone design, especially while away from home (*e.g.,* while shopping). However, all participants found the Attached to Phone design to be more difficult to use than the others for reading longer passages in its current implementation, suggesting that more robust motion tracking and image stabilization are needed to improve the reading experience.

Smartphone Camera. The two participants who had used the previous prototype (P1 and P3) felt that the updated design was an improvement, with a better camera and more usable interactions. However, while the previous design was lightweight and could be used hands-free, the updated design required holding the iPhone steady in midair while reading. All participants found this to be somewhat difficult because of the additional physical effort and shaky image due to unsteady hands. This issue was initially exacerbated by a sometimes slow and uneven frame rate streaming the video from the phone to the HoloLens, which we fixed after the first two sessions, but it also prompted us to add the "freeze frame" feature described above. This feature functioned similarly to existing features on smartphone magnifiers, but with a significantly larger virtual display. Later participants (P4, P6, and P7) liked this feature and found it to be much easier to use than live video when reading longer passages. The issue of image stability could also be addressed in the future by including a portable mount to help hold the phone steady, by adding optical or digital image stabilization, or by integrating the camera and motion controls into a smaller design (*e.g.,* a smartwatch).

The HoloLens Device. While replacing the finger-worn camera with an iPhone camera addressed one aspect of the limited field of view from the previous design (allowing more text to be captured at once), the narrow window that the HoloLens could display was still too small for some participants. This issue was particularly problematic for the two participants with central vision loss, one of whom was completely unable to use the system to read (P5) and one of whom was frustrated by how little of his available vision could be used (P1). In contrast, another participant with tunnel vision found the display to be perfectly sized. The contrast of the HoloLens display also continued to be problematic for some participants, especially for P5 who was unable to see anything on the screen without blocking out all external light. These highly variable results re-emphasize the need for customizability.

Summary. Our second prototype system improved on several aspects of the first, with a better camera that could capture a greater amount of text, more polished and robust virtual display options, and efficient controls that allowed users to more easily customize the AR display to fit their needs. Participants were largely positive about our updated design, appreciating the options for customization and noting tradeoffs between the three AR display designs as well as advantages compared to existing technology. The design sessions also helped to identify important features and design elements for future AR magnification aids.

DISCUSSION

We reflect on the implications of our findings, focusing on ways to support a range of users with different visual impairments and a range of situations.

Overall Experience with 3D Augmented Reality

Our design sessions explored a novel AR magnification approach. The ability to display content in 3D space enables new interactions that are not possible with handheld devices or head-mounted 2D displays. For example, participants liked that they could use stationary 3D designs to create and position an arbitrarily large virtual display and then read a full document by turning their head, rather than scanning line by line as with other portable reading aids. Participants also liked how the design allowed them to easily multitask, for example, by turning away from the virtual display to speak with someone, then turning back to continue reading.

However, some participants commented that our approach required more effort to use than other reading aids. These participants preferred the simplicity of designs that fixed the display in 2D in front of their vision (*e.g.*, as in [25]), especially when they are trying to concentrate on the content of what they are reading. Further refinements to our designs and additional time for the participants to practice using the system may have improved their impressions of the system, but it is also possible that more traditional reading aids or simple 2D image enhancements may work better for some situations or users.

Reflections on Head-mounted AR vs. Handheld Tools

AR magnification has potential benefits compared to other magnification approaches, but also limitations that must be overcome to offer a compelling alternative to existing aids.

Usability. Smartphone magnifiers are portable and readily available but have limited screen size. Users can hold the phone close to their face to compensate, but that may be uncomfortable for extended periods. Current HMDs do not yet address these issues, but we expect that future iterations will be lighter-weight and provide a perceptually larger display. These physical advances should allow users to read more naturally than with a handheld magnifier.

Flexibility. Our approach separates the camera from the display, allowing users to find a comfortable reading position regardless of the location of the physical world object, and supports customization so that users can adapt the display to their visual needs or context.

Privacy and Discreet Use. Handheld magnifiers and smartphone apps offer portability but may require the user to hold the device close to their face to read, preventing discreet use. While current HMDs attract attention for other

reasons (unusual, bulky), we expect that future designs will be smaller and less noticeable, and that users will be able to use the magnification aid more privately and discreetly than with a handheld screen—a feature that one participant found particularly appealing.

Ergonomics. Physical strain and fatigue are potential problems for many portable magnification aids [8]. This was also a limitation of our prototypes, but future AR designs could use a smaller wearable camera that can be aimed separately from the display for maximum flexibility and comfort. Participants also noted ergonomic problems with the HoloLens, including weight and eyestrain. These issues are also present to some extent with other head-worn vision enhancement systems. Future HMDs will need to be smaller and more ergonomic with screens designed to support a wide range of vision levels.

Recommended Design and Future Work

Based on the design sessions, we propose design recommendations and key features for assistive AR devices.

HMD. Participants liked our use of a transparent display that did not block out external vision, unlike most existing HMD systems (*e.g.*, [25,27]). Therefore, an ideal system should use an optical see-through HMD, but in a more lightweight form factor than the HoloLens, with a larger field of view to better support users with limited central vision. However, if contrast cannot be sufficiently improved in future optical see-through HMD devices, a video display that blocks out external light may be a better choice for some low vision users (*e.g.*, P5). Future work should explore alternative display options and evaluate their suitability for different users and contexts.

Camera. Participants valued flexibility, comfort, a wide field of view, and hands-free use, suggesting the need for a wearable camera that can be aimed separately from the display. The finger-worn and handheld smartphone cameras that we tested did not meet these criteria, but neither do the head-worn cameras used in most existing commercial HMD systems (e.g., [27,29]). A head-worn camera should allow for movement and optical zoom independent of the headset so that target content can be captured without requiring users to precisely position their head. Separate motion of the camera and head is also crucial for allowing users to move their head to scan virtual content in 3D, an interaction which participants found intuitive and useful. This design would likely require the ability to detect the content a user is pointing toward so that it can be magnified (*e.g.,* similar to the interaction used by OrCam [14]). Future work should explore these camera options in more depth.

Virtual display. AR magnification systems should include multiple display options to support different users and situations. We encountered tradeoffs between designs, such as the ease of use and attention required, ability to multitask, usefulness for different situations (*e.g.*, reading a long document *vs.* products in a store), and technical complexity and robustness. The ability to anchor virtual content in 3D space in the physical world can support a more natural and flexible reading experience compared to existing 2D vision enhancement systems, but is also more complicated to implement and may have a steeper learning curve for users. Future work should investigate new AR designs, such as the ability to place multiple displays with different targets or magnification levels, and an option to automatically display enhanced content directly over the text (*e.g.,* on a page or sign).

Limitations

The HoloLens has a narrow and centrally located field of view (estimated at 30° × 17.5°), which did not work well for some users. The translucent "holograms" that the HoloLens displays are also low-contrast, and colors are distorted by the screen material. Two participants could not use the device to read due to these issues, and most of the others mentioned them as limitations as well. We did not evaluate alternative headsets, although we anticipate that future versions of the HoloLens or similar technology will be able to address these issues. Future work should consult vision experts to better assess design requirements and usability for specific visual impairments. While our design sessions were informative and helped identify important design features for AR magnification aids, our study was not controlled, included a relatively small number of participants (7 total), and did not quantitatively evaluate usability or reading speed and comprehension. Future work should investigate camera positions and virtual display designs in more depth and directly compare AR magnification aids against existing technology.

CONCLUSION

This paper explored novel applications of AR to assist low vision users, applying recent technology that can anchor 3D virtual content in the physical world. We explored the AR magnification design space and presented two prototype systems that we evaluated and refined through iterative design sessions with low vision participants. Participants liked the general concept of a head-worn magnification aid for its improved portability, privacy, and ready availability compared to other magnification aids they had used. Participants also identified advantages to our 3D AR approach compared to handheld magnification tools, including a more natural reading experience and the ability to more easily multitask, but also some disadvantages such as a steeper learning curve and limitations of the particular hardware we used. Through our open-ended design and evaluation sessions, we identified several common themes that should inform the design of future AR vision enhancement aids for low vision users.

ACKNOWLEDGMENTS

We thank our participants, the Maryland State Library for the Blind and Physically Handicapped, Seokbin Kang, Dhruv Jain, and Matt Mauriello. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

REFERENCES

1. Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4: 355–385. https://doi.org/10.1162/pres.1997.6.4.355
2. American Foundation for the Blind. CCTVs/Video Magnifiers. Retrieved from http://www.afb.org/prodBrowseCatResults.aspx?CatID=53
3. American Foundation for the Blind. Low Vision Optical Devices. Retrieved from www.afb.org/ProdBrowseCatResults.aspx?CatID=54
4. G L Goodrich, E B Mehr, and N C Darling. 1980. Parameters in the use of CCTV's and optical aids. *American journal of optometry and physiological optics* 57, 12: 881–92. Retrieved March 27, 2018 from http://www.ncbi.nlm.nih.gov/pubmed/6164297
5. Google. Google Glass. Retrieved from https://www.google.com/glass/start/
6. Joan M. Greenbaum and Morten Kyng. 1991. *Design at work : cooperative design of computer systems*. L. Erlbaum Associates. Retrieved April 16, 2018 from https://dl.acm.org/citation.cfm?id=125470
7. R Harper, L Culham, and C Dickinson. 1999. Head mounted video magnification devices for low vision rehabilitation: a comparison with existing technology. *The British journal of ophthalmology* 83, 4: 495–500. https://doi.org/10.1136/BJO.83.4.495
8. ChandraM. Harrison. 2004. Low-vision reading aids: reading as a pleasurable experience. *Personal and Ubiquitous Computing* 8, 3–4: 213–220. https://doi.org/10.1007/s00779-004-0280-0
9. Alex D Hwang and Eli Peli. 2014. An augmented-reality edge enhancement application for Google Glass. *Optometry and vision science : official publication of the American Academy of Optometry* 91, 8: 1021–30. https://doi.org/10.1097/OPX.0000000000000326
10. Shaun K Kane, Chandrika Jayant, Jacob O Wobbrock, and Richard E Ladner. 2009. Freedom to roam. In *Proc. ASSETS 2009*, 115. https://doi.org/10.1145/1639642.1639663
11. Sreekar Krishna, Greg Little, John Black, and Sethuraman Panchanathan. 2005. A wearable face recognition system for individuals with visual impairments. In *Proc. ASSETS 2005*, 106. https://doi.org/10.1145/1090785.1090806
12. R W Massof and D L Rickman. 1992. Obstacles encountered in the development of the low vision enhancement system. *Optometry and vision science : official publication of the American Academy of Optometry* 69, 1: 32–41. Retrieved March 26, 2018 from http://www.ncbi.nlm.nih.gov/pubmed/1371334
13. Suranga Nanayakkara, Roy Shilkrot, Kian Peen Yeo, and Pattie Maes. 2013. EyeRing: A Finger-worn Input Device for Seamless Interactions with Our Surroundings. In *Proc. AH 2013*, 13–20. https://doi.org/10.1145/2459236.2459240
14. OrCam Technologies Ltd. OrCam - See for Yourself. Retrieved from http://www.orcam.com/
15. Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K. Kane. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. In *Proc. CHI 2016*, 4884–4895. https://doi.org/10.1145/2858036.2858130
16. Shrinivas Pundlik, Huaqi Yi, Rui Liu, Eli Peli, and Gang Luo. 2016. Magnifying Smartphone Screen using Google Glass for Low-Vision Users. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*: 1–1. https://doi.org/10.1109/TNSRE.2016.2546062
17. Douglas. Schuler and Aki. Namioka. 1993. *Participatory design : principles and practices*. L. Erlbaum Associates. Retrieved April 16, 2018 from https://dl.acm.org/citation.cfm?id=563076
18. Roy Shilkrot, Jochen Huber, Wong Meng Ee, Pattie Maes, and Suranga Chandima Nanayakkara. 2015. FingerReader. In *Proc. CHI 2015*, 2363–2372. https://doi.org/10.1145/2702123.2702421
19. Laurindo de Sousa Britto Neto, Vanessa Regina Margareth Lima Maike, Fernando Luiz Koch, Maria Cecília Calani Baranauskas, Anderson de Rezende Rocha, and Siome Klein Goldenstein. 2015. A Wearable Face Recognition System Built into a Smartwatch and the Blind and Low Vision Users. In *Proc. ASSETS 2015*, 515–528. https://doi.org/10.1007/978-3-319-29133-8_25
20. Lee Stearns, Victor DeSouza, Jessica Yin, Leah Findlater, and Jon E. Froehlich. 2017. Augmented Reality Magnification for Low Vision Users with the Microsoft Hololens and a Finger-Worn Camera. In *Proc. ASSETS 2017*, 361–362. https://doi.org/10.1145/3132525.3134812
21. Lee Stearns, Ruofei Du, Uran Oh, Catherine Jou, Leah Findlater, David A. Ross, and Jon E. Froehlich. 2016. Evaluating Haptic and Auditory Directional Guidance to Assist Blind People in Reading Printed Text Using Finger-Mounted Cameras. *ACM Transactions on Accessible Computing* 9, 1: 1–38. https://doi.org/10.1145/2914793
22. Lee Stearns, Uran Oh, Leah Findlater, and Jon E. Froehlich. 2018. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4: 1–23. https://doi.org/10.1145/3161416
23. Gianni Virgili, Ruthy Acosta, Lori L Grover, Sharon A

Bentley, and Giovanni Giacomelli. 2013. Reading aids for adults with low vision. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD003303.pub3

24. Yuhang Zhao, Michele Hu, Shafeka Hashash, and Shiri Azenkot. 2017. Understanding Low Vision People's Visual Perception on Commercial Augmented Reality Glasses. In *Proc. CHI 2017*, 4170–4181. https://doi.org/10.1145/3025453.3025949
25. Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. 2015. ForeSee: A Customizable Head-Mounted Vision Enhancement System for People with Low Vision. In *Proc. ASSETS 2015*, 239–249. https://doi.org/10.1145/2700648.2809865
26. Annuska Zolyomi, Anushree Shukla, and Jaime Snyder. 2017. Technology-Mediated Sight: A Case Study of Early Adopters of a Low Vision Assistive Technology. *Proc. ASSETS 2017*: 220–229. https://doi.org/10.1145/3132525.3132552
27. eSight Glasses. Retrieved from https://www.esighteyewear.com/technology
28. NuEyes Pro. Retrieved from https://nueyes.com/nueyes-pro/
29. IrisVision. Retrieved from https://irisvision.com/
30. Microsoft Hololens. Retrieved from https://www.microsoft.com/en-us/hololens
31. ClearView+ HD. Retrieved from us.optelec.com/products/hdebbu22ufa-clearview-hd-22-inch-tft-ultra-flexible-arm.html
32. DaVinci HD. Retrieved from www.enhancedvision.com/low-vision-product-line/davinci-pro.html
33. Ruby Handheld Video Magnifier. Retrieved from http://www.freedomscientific.com/Products/LowVision/Ruby
34. Optelec Compact+ HD. Retrieved from https://us.optelec.com/products/compact-hd.html
35. Apple iOS Accessibility Features. Retrieved from https://www.apple.com/accessibility/iphone/vision/
36. Google Android Accessibility Features. Retrieved from https://support.google.com/accessibility/android/answer/6006949
37. Microsoft Seeing AI. Retrieved from https://www.microsoft.com/en-us/seeing-ai/
38. Pivothead Seeing AI. Retrieved from http://www.pivothead.com/seeingai/
39. Oculus Rift. Retrieved from https://www.oculus.com/rift/
40. Apple iOS ARKit. Retrieved from https://developer.apple.com/arkit/

Applying Transfer Learning to Recognize Clothing Patterns Using a Finger-Mounted Camera

Lee Stearns[1], Leah Findlater[2], Jon E. Froehlich[2]
[1]University of Maryland, College Park; [2]University of Washington, Seattle
lstearns@umd.edu; {leahkf, jfroehli}@uw.edu

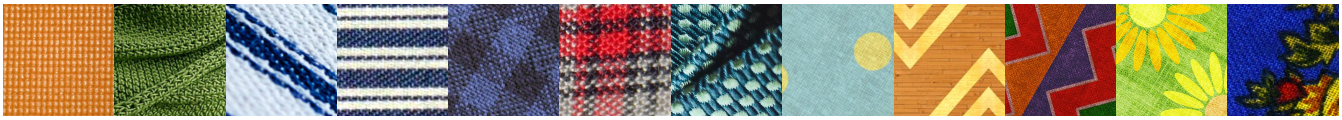


Figure 1. Examples of the six classes in our fabric pattern dataset: solid, striped, checkered, dotted, zigzag, and floral.

ABSTRACT
Color identification tools do not identify visual patterns or allow users to quickly inspect multiple locations, which are both important for identifying clothing. We are exploring the use of a finger-based camera that allows users to query clothing colors and patterns by touch. Previously, we demonstrated the feasibility of this approach using a small, highly-controlled dataset and combining two image classification techniques commonly used for object recognition. Here, to improve scalability and robustness, we collect a dataset of fabric images from online sources and apply transfer learning to train an end-to-end deep neural network to recognize visual patterns. This new approach achieves 92% accuracy in a general case and 97% when tuned for images from a finger-mounted camera.

Author Keywords
Visually impaired users; wearables; pattern recognition

ACM Classification Keywords
Human-centered computing → Accessibility technologies

INTRODUCTION
While color identification tools for users with visual impairments are widely available (e.g., [1,5]), they do not identify visual patterns or allow users to quickly inspect multiple locations—both of which are important for recognizing clothing [2]. Showing promise for more advanced clothing pattern identification, Yuan *et al.* [12–14] developed a system to identify 4 patterns and 11 colors in images captured with a mobile phone or head-mounted camera. Blind users responded positively to the system, although more detailed identification was desired. The interaction was also inefficient, requiring the user to hold out the clothing in front of them and use speech input to individually capture each still image to be classified.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ASSETS '18, October 22–24, 2018, Galway, Ireland
© 2018 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-5650-3/18/10.
https://doi.org/10.1145/3234695.3241015

In contrast, we are investigating a finger-mounted camera approach that allows users to move their finger across an article of clothing, combining tactile information with continuous audio description of the fabric's appearance (Figure 2). In previous work [7], we captured 520 images across 29 articles of clothing using this finger-mounted approach. We combined two complementary image features commonly used for object recognition [3] to train a support vector machine, achieving high classification accuracy (99.4%). While useful for demonstrating feasibility, the dataset was highly controlled, which risks overfitting, and the training process was not easily scalable.

In this paper, we build a larger and more varied dataset of images from online sources (Figure 1), which should allow our system to identify previously *unseen* fabrics—for example, to support shopping for new clothes. To assess whether this Internet-based dataset can be used to identify patterns in images collected with our finger-mounted system, we apply transfer learning to train an end-to-end deep neural network and test with the previously collected finger-mounted camera images. Our contributions include: (i) a novel dataset of 77,052 fabric images across six visual patterns; (ii) results of applying transfer learning to train a convolutional neural network for classifying fabric patterns; (iii) a preliminary system to describe clothing appearance in realtime using images from a finger-mounted camera. Because this work is still at an early stage, we also discuss open questions and challenges important for future systems.

DATA COLLECTION AND DATASETS
Existing texture datasets include textures that can easily be distinguished by touch or that are not relevant to clothing. For example, the *Describable Textures Dataset* [3] includes *braided* and *frilly*, and our prior dataset [7] includes *denim*,

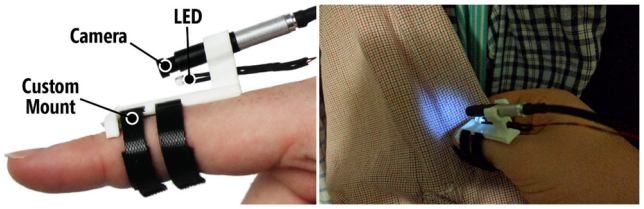


Figure 2. Prototype wearable camera and LED system, the same used in previous work [7]: (a) Close-up view of system, (b) Identifying an article of clothing.

knitted, and *lacelike*. While automatic identification of these textures may be useful to avoid misclassifications, in general they are not necessary to assist blind users. Instead, we selected six common visual patterns that are difficult or impossible to distinguish by touch alone: *solid*, *striped*, *checkered*, *dotted*, *zigzag*, and *floral* (Figure 1).

To create our dataset, we added the word "fabric" after each class name and downloaded the top 1000 search results from Google Images.[1] After one author manually removed erroneous results and cropped others as necessary (*e.g.*, to remove logos or background imagery), the dataset contained between 317 and 584 images per class (2764 images total). We augmented this data using a standard process to increase the training set size and improve robustness [10], rotating each image in 30-degree increments and cropping the center at multiple scales (1–4 depending on the resolution of the original image), resulting in 8232–17,304 samples per class, or 77,052 images total[2].

ALGORITHMS AND VALIDATION

To identify textures, we repurposed a state-of-the-art convolutional neural network model (ResNet-101 [6]) that was pre-trained on the *ImageNet* object dataset [9]. Using a standard transfer learning approach to avoid overfitting when insufficient data is available [4], we fixed all layers except for the final densely connected classification layer, and trained the weights for that layer using our dataset. To ensure that each class was equally likely when training, we randomly sampled 6400 images per class for the training set and 1600 images for a test set, discarding the rest.

Classification accuracy on the test set was 91.7%, suggesting that this approach should work well in general. On our smaller finger-mounted camera dataset [7], which contained 400 images across the six classes and included variations in distance, rotation, perspective, and fabric tension, accuracy was 72.8%. Most errors were caused by confusion due to insufficient context or coarse threads (*e.g.*, Figure 3). For example, *zigzag* was the worst performing class, likely because the camera's proximity to the fabric obscured much of the pattern. Roughly 14.5% of images were also misclassified as *checkered* or *floral*, most likely due to confusion from the coarse threads. Half of the images were captured with the finger-worn camera held 5cm from the fabric, while the other half were captured from a distance of 12cm; if only the latter images are considered (N=200), accuracy rises to 78.0%. Finally, continuing the training process to fine-tune the classifier using roughly half of the finger-camera images (N=36 per class) increases accuracy to 96.5%, suggesting additional images from the target domain can boost performance.

ONGOING WORK

In addition to identifying patterns, our real-time system identifies the dominant color in each image. To do so, we use k-means clustering with a dynamic number of clusters selected using the "jump method" [11], which searches for the point at which additional clusters stop greatly reducing error. We name each cluster center using the XKCD color survey results [8] to provide commonly accepted names for 48 RGB color values. The level of detail is user-configurable, including the number and complexity of the colors named in each image (*e.g.*, "green, purple, brown", or "lime green, lilac, beige"). Users can also identify multiple colors by moving their finger across the fabric. We did not evaluate color accuracy but if properly calibrated, our finger-worn approach should mitigate issues with lighting and distance that impact existing color identifiers.

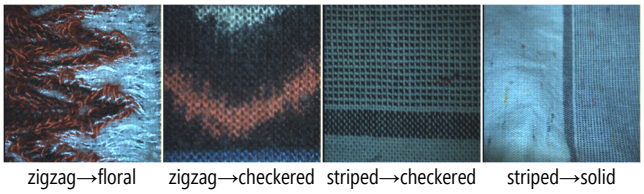


Figure 3. Example misclassifications (actual→predicted)

The system runs at approximately four frames per second on a desktop computer.[3] The current implementation tracks classification results for the most recent two seconds. To reduce noise from misclassifications, patterns are reported by majority vote, with unclear results labeled "unknown". Colors are only reported if they are named consistently across frames—for example, results of "blue and light blue" and "blue and gray" would be reported simply as "blue". Users can press a button to hear the most recent result via text to speech, or hold for continuous updates.

DISCUSSION AND FUTURE WORK

Our pattern classification accuracy when fine-tuned with images from a finger-mounted camera is similar to our prior work (97% vs 99%), but our end-to-end deep learning approach reduces training overhead and should be much more generalizable and robust. Our use of a single online source does introduce some bias in our dataset, which future work should investigate. And to mitigate errors caused by lack of context or distracting details (*e.g.*, coarse threads), the camera should likely be positioned farther back on the user's finger or wrist. This change would still allow users to easily query multiple locations and combine automated feedback with their own sense of touch.

Future work will need to investigate performance and usability with visually impaired users, and assess the potential benefits of our approach compared to existing aids. Additionally, how best to convey complex color and pattern information to users is still an open question.

ACKNOWLEDGMENTS

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs under Award W81XWH-14-1-0617.

[1] https://github.com/hardikvasa/google-images-download
[2] https://github.com/lstearns86/clothing-pattern-dataset
[3] Dell Precision Workstation, dual Intel Xeon CPU @ 2.1 GHz, NVIDIA GeForce GTX 1080

REFERENCES

1. Brytech. Brytech Color Teller. Retrieved from http://www.brytech.com/colorteller/
2. Michele A Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P Bigham, and Amy Hurst. 2012. Crowdsourcing Subjective Fashion Advice Using VizWiz: Challenges and Opportunities. In *Proceedings of ASSETS 2012*, 135–142. https://doi.org/10.1145/2384916.2384941
3. Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. 2014. Describing Textures in the Wild. In *Proceedings of CVPR 2014*, 3606–3613.
4. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. DeCAF: a deep convolutional activation feature for generic visual recognition. In *Proceedings of ICML 2014*, I-647. Retrieved June 13, 2018 from https://dl.acm.org/citation.cfm?id=3044879
5. GreenGar Studios. Color Identifier. Retrieved from https://itunes.apple.com/us/app/color-identifier/id363346987?mt=8
6. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proceedings of CVPR 2016*, 770–778. https://doi.org/10.1109/CVPR.2016.90
7. Alexander J. Medeiros, Lee Stearns, Leah Findlater, Chuan Chen, and Jon E. Froehlich. 2017. Recognizing Clothing Colors and Visual Textures Using a Finger-Mounted Camera: An Initial Investigation. In *Proceedings of ASSETS 2017*, 393–394. https://doi.org/10.1145/3132525.3134805
8. Randall Munroe. 2010. Color Survey Results. Retrieved from https://blog.xkcd.com/2010/05/03/color-survey-results/
9. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 3: 211–252. https://doi.org/10.1007/s11263-015-0816-y
10. Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. Retrieved June 12, 2018 from http://arxiv.org/abs/1409.1556
11. Catherine A Sugar and Gareth M James. 2003. Finding the Number of Clusters in a Dataset. *Journal of the American Statistical Association* 98, 463: 750–763. https://doi.org/10.1198/016214503000000666
12. Yingli Tian and Shuai Yuan. 2010. Clothes Matching for Blind and Color Blind People. In *Proceedings of ICCHP 2010*, 324–331. https://doi.org/10.1007/978-3-642-14100-3_48
13. Xiaodong Yang, Shuai Yuan, and YingLi Tian. 2014. Assistive Clothing Pattern Recognition for Visually Impaired People. *IEEE Transactions on Human-Machine Systems* 44, 2: 234–243. https://doi.org/10.1109/THMS.2014.2302814
14. Shuai Yuan, YingLi Tian, and Aries Arditi. 2011. Clothing matching for visually impaired persons. *Technology and Disability* 23, 2: 75–85. https://doi.org/10.3233/TAD-2011-0313