FingerReader: A Wearable Device to Explore Printed Text on the Go

FingerReader: A Wearable Device to Explore Printed Text on the Go

Roy Shilkrot[1*], **Jochen Huber**[1,2*], **Meng Ee Wong**[3], **Pattie Maes**[1], **Suranga Nanayakkara**[2]
[1] MIT Media Lab, Cambridge, MA USA. {roys,jhuber,pattie}@mit.edu
[2] Singapore University of Technology and Design, Singapore. suranga@sutd.edu.sg
[3] National Institute of Education, Nanyang Technological University, Singapore. mengee.wong@nie.edu.sg

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

Author Keywords
Assistive technology; Text reading; Wearable interface;

ACM Classification Keywords
K.4.2 Social Issues: Assistive technologies for persons with disabilities; B.4.2 Input/Output Devices: Voice

INTRODUCTION
Some people with a visual impairment (VI) find it difficult to access text documents in different situations, such as reading text on the go and accessing text in non-ideal conditions (e.g. low lighting, unique layout, non-perpendicular page orientations), as reported in interviews we conducted with assistive technology users. We found that available technologies, such as smartphone applications, screen readers, flatbed scanners, e-Book readers, and embossers, are considered to have slow processing speeds, poor accuracy or cumbersome usability. Day-to-day text-based information, such as bus and train station information boards, are said to be generally inaccessible, which greatly affects the mobility and freedom of people with a VI outside the home, the Royal National Institute of Blind People (RNIB) reports [13]. Technological barriers inhibit blind people's abilities to gain more independence, a characteristic widely identified as important by our interviewees.

[*]These authors contributed equally to this work.

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea.

ACM 978-1-4503-3145-6/15/04$15.00.
http://dx.doi.org/10.1145/2702123.2702421

In this paper, we present our work from the past 18 months of creating a mobile device to tackle some of the problems current text reading technologies present to blind users. Our work contributes to the growing pool of assistive reading devices in three primary ways:

- First, we share the results of interview sessions with blind users that uncover problems with existing text reading solutions, as well as expectations for future assistive devices and their capabilities. Our design choices are based on these findings.
- Second, we conceptualize and implement FingerReader, a finger-worn system for local-sequential text scanning, where the user scans the text progressively in a local view and hears the recognized words synthesized to audible speech. It enables continuous feedback to the user and allows for new ways of reading, such as non-linear skimming to different parts of the text. Our proposed method utilizes computer vision algorithms, along with audio and tactile cues for effectively guiding the user in reading printed text using the fingertip as a cursor.
- Last, we report findings from three evaluations: a technical evaluation to understand the text extraction accuracy, user feedback sessions with blind participants to assess the feedback mechanism, and an end-to-end study to assess the system's real-world feasibility and explore further design opportunities.

RELATED WORK
Researchers in both academia and industry exhibited a keen interest in aiding people with VI to read printed text. The earliest evidence we found for a specialized assistive text-reading device for the blind is the Optophone, dating back to 1914 [3]. However the Optacon [10], a steerable miniature camera that controls a tactile display, is a more widely known device from the mid 20th century. Table 1 presents more contemporary methods of text-reading for the VI based on key features: adaptation for non-perfect imaging, type of text, User Interface (UI) suitable for VI and the evaluation method. Thereafter we discuss related work in three categories: wearable devices, handheld devices and readily available products.

As our device is finger worn, we refer the reader to our prior work that presents much of the related finger worn devices [12], and to [16] for a limited survey of finger-worn devices for general public use. Additionally, we refer readers to the encompassing survey by Lévesque [9] for insight into the use of tactiles in assistive technology.

Publication	Year	Interface	Type of Text	Response Time	Adaptation	Evaluation	Reported Accuracy
Ezaki et al. [4]	2004	PDA	Signage			ICDAR 2003	P 0.56 R 0.70
Mattar et al. [11]	2005	Head-worn	Signage		Color, Clutter	Dataset	P ?.?? R 0.90[1]
Hanif and Prevost [5]	2007	Glasses, Tactile	Signage	43-196s		ICDAR 2003	P 0.71 R 0.64
SYPOLE [15]	2007	PDA	Products, Book cover	10-30s	Warping, Lighting	VI users	P 0.98 R 0.90[1]
Pazio et al. [14]	2007		Signage		Slanted text	ICDAR 2003	
Yi and Tian [24]	2012	Glasses	Signage, Products	1.5s	Coloring	VI users	P 0.68 R 0.54
Shen and Coughlan [18]	2012	PDA, Tactile	Signage	<1s		VI users	
Kane et al. [7]	2013	Stationery	Printed page	Interactive	Warping	VI users	
Stearns et al. [23]	2014	Finger-worn	Printed page	Interactive	Warping	VI users	
Shilkrot et al. [19]	2014	Finger-worn	Printed page	Interactive	Slanting, Lighting	VI users	

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

Table 1: Recent efforts in academia of text-reading solutions for the VI. Accuracy is in precision (P) recall (R) form, as reported by the authors.

Wearable devices

In a wearable form-factor, it is possible to use the body as a directing and focusing mechanism, relying on proprioception or the sense of touch, which are of utmost importance for people with VI. Yi and Tian [24] placed a camera on shade-glasses to recognize and synthesize text written on objects in front of them, and Hanif and Prevost's [5] did the same while adding a handheld device for tactile cues. Mattar et al. are using a head-worn camera [11], while Ezaki et al. developed a shoulder-mountable camera paired with a PDA [4]. Differing from these systems, we proposed using the finger as a guide [12], and supporting sequential acquisition of text rather than reading text blocks [19]. This concept has inspired other researchers in the community [23].

Handheld and mobile devices

Mancas-Thillou, Gaudissart, Peters and Ferreira's SYPOLE consisted of a camera phone/PDA to recognize banknotes, barcodes and labels on various objects [15], and Shen and Coughlan recently presented a smartphone based sign reader that incorporates tactile vibration cues to help keep the text-region aligned [18]. The VizWiz mobile assistive application takes a different approach by offloading the computation to humans, although it enables far more complex features than simply reading text, it lacks real time response [1].

Assistive mobile text reading products

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

Prototypes and products in all three categories, save for [23], follow the assumption that the goal is to consume an entire block of text at once, therefore requiring to image the text from a distance or use a special stand. In contrast, we focused on creating a smaller and less conspicuous device, allowing for intimate operation with the finger that will not seem strange to an outside onlooker, following the conclusions of Shinohara and Wobbrock [21]. Giving the option to read locally, skim over the text at will in a varying pace, while still being able to read it through, we sought to create a more liberating reading experience.

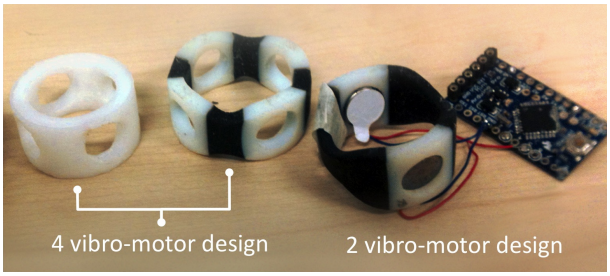
FOCUS GROUP SESSIONS

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

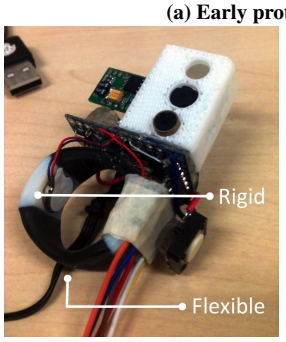
- All participants routinely used flatbed scanners and camera-equipped smartphones to access printed text.
- While flatbed scanners were reported to be easy to use, participants mentioned problems when scanning oddly shaped prints. Our participants preferred mobile devices due to their handiness, but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

Following the findings from the focus group sessions, we set to design a device that enables: skimming through the text,

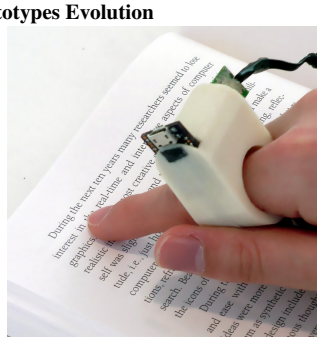
[1] `http://www.knfbreader.com`
[2] `http://blindsight.com`
[3] `http://www.abbyy.com/textgrabber`
[4] `http://standscan.com`
[5] `http://www.docscannerapp.com/saytext`
[6] `http://mobile.aisquared.com`
[7] `http://www.creaceed.com/iprizmo`
[8] `http://www.seeingwithsound.com`
[9] `http://www.abisee.com`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



(a) Early prototypes Evolution



(b) Multi-material prototype



(c) New prototype

Figure 1: FingerReader prototypes.

have a real time single-handed operation, and provides multimodal continuous feedback.

FINGERREADER: A WEARABLE READING DEVICE

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

The design of the FingerReader is a continuation of our work on finger wearable devices for seamless interaction [12, 19], and inspired by the focus group sessions. Exploring the design concepts with blind users revealed the need to have a small, portable device that supports free movement, requires minimal setup and utilizes real-time, distinctive multimodal response. The finger-worn design keeps the camera in a fixed distance from the text and utilizes the inherent finger's sense of touch when scanning text on the surface. Additionally, the device provides a simple interface for users as it has no buttons, and affords to easily identify the side with the camera lens for proper orientation.

Hardware Details

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

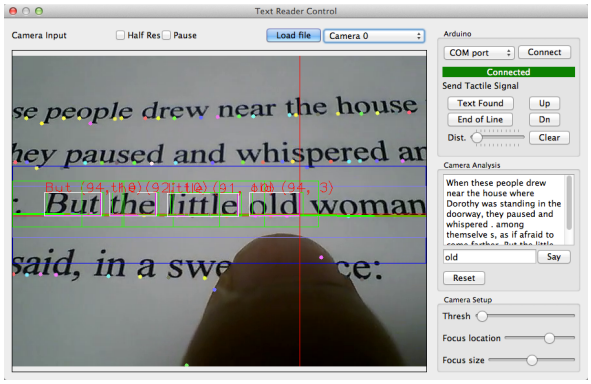


Figure 2: Our software in midst of reading, showing the detected line, words and the accumulated extracted text

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

Algorithms and Software

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

Vision Algorithm Overview

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

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

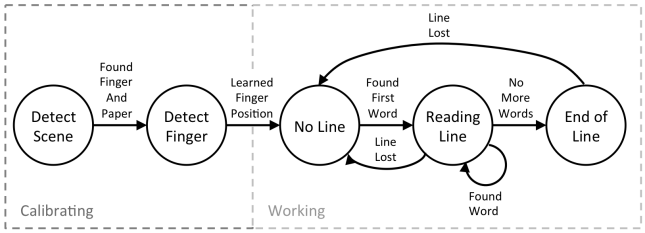


Figure 3: Sequential text reading algorithm state machine.

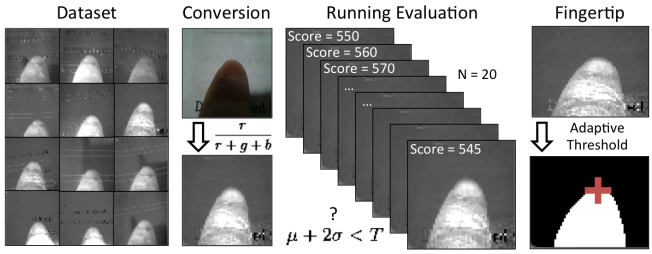


Figure 4: Scene and fingertip detection.

Scene and Finger Detection: The initial calibration step tries to ascertain whether the camera sees a finger on a contrasting paper. The input camera image is converted to the normalized-RGB space: $(R,G,B) = \left(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b}\right)$, however we keep only the normalized red channel (R) that corresponds well with skin colors and ameliorates lighting effects. The monochromatic image is downscaled to 50x50 pixels and matched to a dataset of prerecorded typical images of fingers and papers from a proper perspective of the device camera. To score an incoming example image, we perform a nearest neighbor matching and use the distance to the closest database neighbor. Once a stable low score is achieved (by means of a running-window of 20 samples and testing if $\mu_{score} + 2\sigma < threshold$) the system deems the scene to be a well-placed finger on a paper, issues an audio command and advances the state machine. See Fig. 4 for an illustration of this process.

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

The inlying fingertip detection guides a local horizontal *focus region*, located above the fingertip, within which the following states perform their operations. The focus region helps with efficiency in calculation and also reduces confusion for the line extraction algorithm with neighboring lines (see Fig. 5). The height of the focus region may be adjusted as a parameter, but the system automatically determines it once a text line is found.

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

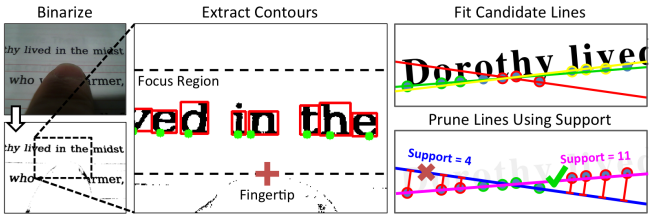


Figure 5: Text line extraction process.

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

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

Word Tracking and Signaling: Whenever a new word is recognized it is added to a pool of words to track along with its initial bounding rectangle. For tracking we use template matching, utilizing image patches of the words and an L_2-norm matching score. Every successful tracking, marked by a low matching score and a feasible tracking velocity (i.e. it corresponds with the predicted finger velocity for that frame), contributes to the bank of patches for that word as well as to the prediction of finger velocity for the next tracking cycle. To maintain an efficient tracking, we do not search the entire frame but constrain the search region around the last position of the word while considering the predicted movement speed. We also look out for blurry patches, caused by rapid movement and the camera's rolling shutter, by binarizing the patch and counting the number of black vs. white pixels. A ratio of less than 25% black is considered a bad patch to be discarded. If a word was not tracked properly for a set number of frames

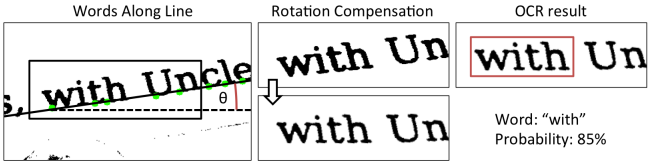


Figure 6: Word extraction process.

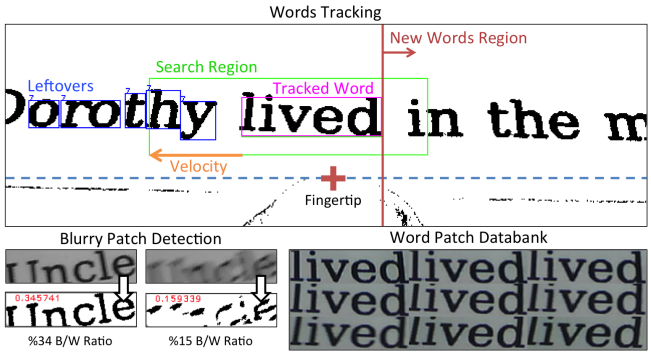


Figure 7: Word tracking process.

we deem as "lost", and remove it from the pool. See Fig. 7 for an illustration.

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

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

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

EVALUATION

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

Across all studies we were only able to recruit a small number of participants. In accordance with Sears and Hanson, who attest to the difficulty of recruiting participants with similar impairment condition in accessibility research [17]. Thus we looked to maximize the results we can obtain from a small number of participants with different impairment histories instead of striving for generalizability (sensu [20]).

Technical Accuracy Analysis

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

The accuracy was defined as as $acc = 1 - LD_{norm}$, with LD_{norm} being the normalized Levenshtein Distance (LD) [8] between the scanned and original text. The LD counts the number of character edits between two strings, e.g. $LD(\text{"hello"}, \text{"h3ll0"}) = 2$; a higher distance means a less accurate scan. To normalize, we divided the LD by the number of characters in the paragraph (either scanned or original) with the maximal number of characters: $LD_{norm} = LD/\max(S_{scan}, S_{orig})$, where S_i is the length of scanned string i (the scanned string can be larger than the original).

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

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

User Feedback on Cueing Modalities

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

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

how FingerReader can help them conduct daily tasks, and be able to explore printed text in their surroundings in a novel way.

Print Access and Reading Study

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

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

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

Participants and Study Design

Following the approach of Sears and Hanson [17], we hereby detail the participants information. All participants were blind, P2 and P3 since birth and consequently have never experienced text visually (see table 2). P1 became blind at the age of 18. Before that, he considered himself an avid reader. P2 has very low light perception, P3 no light perception at all. All participants had perfect hearing and were right-handed. All participants had prior exposure to the FingerReader, which included brief demonstrations during recruitment to make sure participants are comfortable before committing to the study.

They all share stationary text access habits, e.g. in using a screenreader like JAWS to access digital text on a PC or Mac or in scanning printed documents to have them read back e.g. with ABBYY FineReader. On the go, P1 and P2 mostly rely on the help of sighted people to read relevant text to them. Specifically, P2 has never owned a smartphone and does not consider himself tech-savvy. Both P1 and P3 own an iPhone and use it to access digital information using Apple's VoiceOver technology. Yet, P2 considers himself only an occasional user of technology. P3 was the most tech-savvy participant. He regularly uses mobile applications on his iPhone to access printed text on the go, namely TextGrabber and Prizmo. P3 stated that he uses either software as a backup in case the other fails to detect text properly. He described himself as an avid user of a foldable StandScan, yet he seldom carries it with him as it is too 'bulky and cumbersome'. In mobile settings, he usually captures documents free-handedly by applying a two-step process where he first places the print in landscape and centers the iPhone on top of it (*framing*) and then lifts the iPhone chin-high to take a picture and have the software read the text back to him (*capture*). The whole capturing process takes him on average 2.5 minutes, excluding any trial and error and without having the text read back to him.

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

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

Table 2: Overview of the participants from the text exploration study.

Method and Tasks

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

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

Session 2: The second sessions included a short practice session to let participants refamiliarize themselves with the device. This was followed by a repetition of the text access tasks to qualitatively compare their performance to the first session. Next, each participant was given a set of 5 articles taken from the online edition of a local newspaper (see Fig. 8). All articles were set in a single-column layout and did not contain pictures. Each participant was asked to explore the news articles and report the gist of the article. The sessions were concluded with a questionnaire (inspired by [6]).

Each set of tasks was fully counterbalanced, and all feedback modalities were available. As for data gathering techniques, we video-recorded the interactions, lead semi-structured interviews, observed the participants during the session and

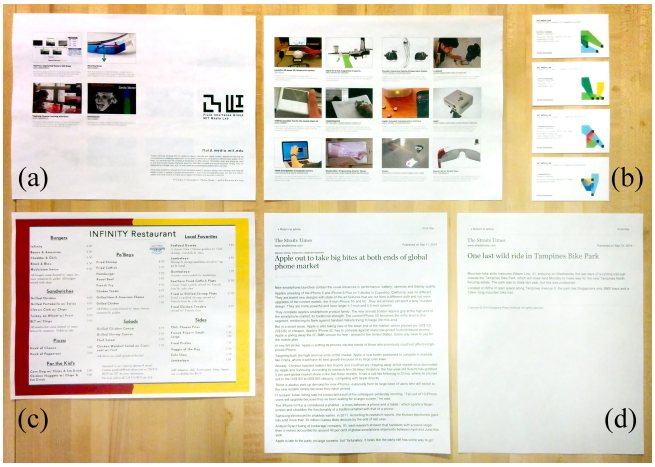


Figure 8: Documents used in the study: (a) pamphlet, (b) business cards, (c) restaurant menu and (d) newspaper articles. The font size varied across documents: 10-14pt (business cards, pamphlet), 12-24pt (menu), 12-14pt (newspaper articles).

asked the participants to think aloud. Two authors acted as experimenters in both sessions. One experimenter had perfect vision while the other was blind with only little peripheral light perception (since age 10).

Results

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

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

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

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

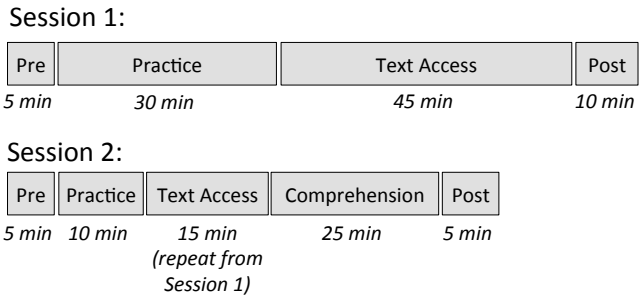


Figure 9: Overview of the schedule for both sessions. "Pre" and "Post" refer to interviews held before and after the tasks.

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

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

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

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

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

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

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

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

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

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

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

Table 3: Results from the questionnaire on 5-point Likert scale (1=stronlgy disagree, 5=strongly agree). The comparison to other mobile text reading aids was only applicable to P3.

- Visual layout: The restaurant menu and the business contained were typeset in different layouts, e.g. a multi-column one. This was particularly challenging for the participants, less so for P1. P2 and P3 were specifically challenged by the multi-column layouts, as e.g. "formattings do not exist" (P2).
- Synthesized voice: The voice from the employed text-to-speech engine was difficult to understand at times. Particularly P3 mentioned that it was hard for him to distinguish the voice from the audio feedback and thus missed a bunch of words occasionally. This lead him to employ the re-reading technique mentioned above.
- Audio feedback: P1 and P2 preferred the audio feedback over the tactile feedback and wished for an audio-only mode. P2 mentioned that the choice of audio feedback could be better, as he found it hard to distinguish high-pitch tones (line deviation) from low-pitch tones (finger twisting/rotation) which he called the "High-Low-Orchestra".
- Fatigue: All participants reported that they would not use the FingerReader for longer reading sessions such as books, as it is too tiring. In this case, they would simply prefer an audio book or a scanned PDF that is read back, e.g. using ABBYY FineReader (P1).
- Serendipity: Whenever any of the participants made the FingerReader read the very first correct word of the document, they smiled, laughed or showed other forms of excitement–every single time. P2 once said that is an "eye-opener". P1 said that it is "encouraging".

Table 3 shows the results for the post-study questionnaire from session 2. The overall experience with the FingerReader was rated as mediocre by all participants. They commented that this was mainly due to the synthesized voice being unpleasent and the steep learning curve in session 1, with session 2 being less difficult (cf. comments above).

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

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

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

DISCUSSION

We discuss the results from the evaluation in the following and highlight lessons learned from the development of the FingerReader. We hope that these insights will help other

researchers in the field of finger-worn reading devices for the blind and inform the design of future devices.

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

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

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

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

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

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

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

LIMITATIONS

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

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

CONCLUSION

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

First, our observations suggest that a local-sequential approach is beneficial for document exploration–but not as

much for longer reading sessions, due to troublesome navigation in complex layouts and fatigue. Access to small bits of text, as found on business cards, pamphlets and even newspaper articles, was considered viable. Second, we observed a rich set of interaction strategies that shed light onto potential real-world usage of finger-worn reading aids. A particularly important insight is the direct correlation between the finger movement and the output of the synthesized speech: navigating within the text is closely coupled to navigating in the produced audio stream. Our findings suggest that a direct mapping could greatly improve interaction (e.g. easy "re-reading"), as well as scaffold the mental model of a text document effectively, avoiding "ghost text". Last, although our focus sessions on the feedback modalities concluded with an agreement for cross-modality, the thorough observation in the follow-up study showed that user preferences were highly diverse. Thus, we hypothesize that a *universal* finger-worn reading device that works uniformly across all users may not exist (sensu [20]) and that personalized feedback mechanisms are key to address needs of different blind users.

In conclusion, we hope the lessons learned from our 18-month-long work on the FingerReader will help peers in the field to inform future designs of finger-worn reading aids for the blind. The next steps in validating the FingerReader are to perform longer-term studies with specific user groups (depending on their impairment, e.g. congenitally blind, late-blind, low-vision), investigate how they appropriate the FingerReader and derive situated meanings from their usage of it. We also look to go beyond usage for persons with a visual impairment, and speculate the FingerReader may be useful to scaffold dyslexic readers, support early language learning for preschool children and reading non-textual languages.

ACKNOWLEDGMENTS

We thank C. Liu for help with the user studies and product design. We thank the MIT Media Lab Consortia and the MIT-SUTD International Design Center for funding this work. Finally we thank all study participants for their time and valuable remarks.

REFERENCES

1. Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S., and Yeh, T. Vizwiz: Nearly real-time answers to visual questions. In *Proc. of UIST*, ACM (2010), 333–342.
2. Black, A. W., and Lenzo, K. A. Flite: a small fast run-time synthesis engine. In *4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis* (2001).
3. d'Albe, E. F. On a type-reading optophone. *Proceedings of the Royal Society of London. Series A 90*, 619 (1914), 373–375.
4. Ezaki, N., Bulacu, M., and Schomaker, L. Text detection from natural scene images: towards a system for visually impaired persons. In *Proc. of ICPR*, vol. 2 (2004), 683–686.
5. Hanif, S. M., and Prevost, L. Texture based text detection in natural scene images-a help to blind and visually impaired persons. In *CVHI* (2007).
6. Harada, S., Sato, D., Adams, D. W., Kurniawan, S., Takagi, H., and Asakawa, C. Accessible photo album: Enhancing the photo sharing experience for people with visual impairment. In *Proc. of CHI*, ACM (2013), 2127–2136.
7. Kane, S. K., Frey, B., and Wobbrock, J. O. Access lens: a gesture-based screen reader for real-world documents. In *Proc. of CHI*, ACM (2013), 347–350.
8. Levenshtein, V. I. Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, vol. 10 (1966), 707.
9. Lévesque, V. Blindness, technology and haptics. *Center for Intelligent Machines* (2005), 19–21.
10. Linvill, J. G., and Bliss, J. C. A direct translation reading aid for the blind. *Proc. of the IEEE 54*, 1 (1966), 40–51.
11. Mattar, M., Hanson, A., and Learned-Miller, E. Sign classification using local and meta-features. In *CVPR - Workshops*, IEEE (June 2005), 26–26.
12. Nanayakkara, S., Shilkrot, R., Yeo, K. P., and Maes, P. EyeRing: a finger-worn input device for seamless interactions with our surroundings. In *Proc. of Augmented Human* (2013), 13–20.
13. Pavey, S., Dodgson, A., Douglas, G., and Clements, B. Travel, transport, and mobility of people who are blind and partially sighted in the uk. Royal National Institute for the Blind (RNIB), April 2009.
14. Pazio, M., Niedzwiecki, M., Kowalik, R., and Lebiedz, J. Text detection system for the blind. In *15th european signal processing conference EUSIPCO* (2007), 272–276.
15. Peters, J.-P., Thillou, C., and Ferreira, S. Embedded reading device for blind people: a user-centered design. In *Proc. of ISIT*, IEEE (2004), 217–222.
16. Rissanen, M. J., Vu, S., Fernando, O. N. N., Pang, N., and Foo, S. Subtle, natural and socially acceptable interaction techniques for Ringterfaces-Finger-Ring shaped user interfaces. In *Distributed, Ambient, and Pervasive Interactions*. Springer, 2013, 52–61.
17. Sears, A., and Hanson, V. Representing users in accessibility research. In *Proc. of CHI*, ACM (2011), 2235–2238.
18. Shen, H., and Coughlan, J. M. Towards a real-time system for finding and reading signs for visually impaired users. In *Proc. of ICCHP*, Springer (2012), 41–47.
19. Shilkrot, R., Huber, J., Liu, C., Maes, P., and Nanayakkara, S. C. Fingerreader: A wearable device to support text reading on the go. In *CHI EA*, ACM (2014), 2359–2364.
20. Shinohara, K., and Tenenberg, J. A blind person's interactions with technology. *Commun. ACM 52*, 8 (Aug. 2009), 58–66.
21. Shinohara, K., and Wobbrock, J. O. In the shadow of misperception: Assistive technology use and social interactions. In *Proc. of CHI*, ACM (2011), 705–714.
22. Smith, R. An overview of the tesseract OCR engine. In *Proc. of ICDAR*, vol. 2, IEEE (2007), 629–633.
23. Stearns, L., Du, R., Oh, U., Wang, Y., Findlater, L., Chellappa, R., and Froehlich, J. E. The design and preliminary evaluation of a finger-mounted camera and feedback system to enable reading of printed text for the blind. In *Workshop on Assistive Computer Vision and Robotics, ECCV*, Springer (2014).
24. Yi, C., and Tian, Y. Assistive text reading from complex background for blind persons. In *Camera-Based Document Analysis and Recognition*. Springer, 2012, 15–28.