The Tangible Desktop: A Multimodal Approach to Nonvisual Computing

MARK S. BALDWIN, GILLIAN R. HAYES, and OLIVER L. HAIMSON,
University of California, Irvine
JENNIFER MANKOFF and SCOTT E. HUDSON, Carnegie Mellon University

Audio-only interfaces, facilitated through text-to-speech screen reading software, have been the primary mode of computer interaction for blind and low-vision computer users for more than four decades. During this time, the advances that have made visual interfaces faster and easier to use, from direct manipulation to skeuomorphic design, have not been paralleled in nonvisual computing environments. The screen reader–dependent community is left with no alternatives to engage with our rapidly advancing technological infrastructure. In this article, we describe our efforts to understand the problems that exist with audio-only interfaces. Based on observing screen reader use for 4 months at a computer training school for blind and low-vision adults, we identify three problem areas within audio-only interfaces: ephemerality, linear interaction, and unidirectional communication. We then evaluated a multimodal approach to computer interaction called the Tangible Desktop that addresses these problems by moving semantic information from the auditory to the tactile channel. Our evaluation demonstrated that among novice screen reader users, Tangible Desktop improved task completion times by an average of 6 minutes when compared to traditional audio-only computer systems.

CCS Concepts: • **Human-centered computing → Haptic devices**; **Empirical studies in accessibility**; **Accessibility systems and tools**; *HCI design and evaluation methods*; • **Social and professional topics → People with disabilities**;

Additional Key Words and Phrases: Accessibility, visual impairment, blindness, tangible, haptic, hardware, vibrotactile feedback, assistive technology

ACM Reference format:
Mark S. Baldwin, Gillian R. Hayes, Oliver L. Haimson, Jennifer Mankoff, and Scott E. Hudson. 2017. The Tangible Desktop: A Multimodal Approach to Nonvisual Computing. *ACM Trans. Access. Comput.* 10, 3, Article 9 (August 2017), 28 pages.
https://doi.org/10.1145/3075222

1 INTRODUCTION

Desktop and mobile computing systems have traditionally relied on audio output as the primary interface for blind and low-vision users. Using only audio output provides simplicity of implementation and relative ease in which interactions can be learned. In practice, however, this model fails

---

Authors' addresses: M. S. Baldwin, G. R. Hayes, and O. L. Haimson, Department of Informatics, University of California, Irvine, Irvine, CA 92697; J. Mankoff and S. E. Hudson, Human Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213.

https://doi.org/10.1145/3075222

to match the quality of sighted user experiences. Where sighted individuals benefit from the parallel processing provided by vision, blind and low-vision users must process the same information sequentially through a single auditory modality. It is through this translation from visual to auditory that valuable information (e.g., spatial arrangement and text emphasis) is lost. Additionally, the ephemeral nature of audio prevents information from persisting beyond the moment it is delivered.

Despite these challenges, blind and low-vision users still engage with screen readers to serve a variety of needs, including basic home and work productivity as well as emergent forms of technical engagement. Increasingly, just as with sighted users, blind and low-vision users have become dependent on some form of computer interaction for many aspects of everyday life. These uses can be particularly challenging for nonvisual computer users, because they often include images and interactive elements that do not translate well to audio output. Despite legal requirements and general support for accessibility online [43, 51], a variety of challenges still exist for users who must rely solely on audio output, including system inefficiency, poor user experience, and cognitive overload.

Converting visual information to auditory output through text-to-speech engines has long been the most common method for nonvisual computer interaction. Affordable computers with the ability to produce audio output are readily available, standards are in place [49, 50], and there is already significant infrastructure to teach people to use screen readers through community and school-based programs. On the software side, a developer only needs to ensure that visual elements are semantically defined to make them accessible to screen readers. In practice, this is a relatively straightforward task. Thus, it is perhaps not surprising that advances in the communication of information visually, continue to be redirected as audio output for blind and low-vision users.

Even if screen readers were optimal translations of the computer experiences at the time they were developed (a time of command line interaction), the changes we have witnessed in user experiences over the past few decades have destroyed any possibility of these to be comparable interfaces today [28]. Those dependent on screen readers are simply not benefiting from the massive advances in user experience, particularly the two-dimensional direct manipulation of the graphical user interface (GUI). Embedded within the GUI are visual and spatial cues that allow sighted users to perceive multiple actions simultaneously. These cues enable sighted users to rely on recognition rather than recall [30], providing a significant advantage over those using screen readers.

In this work, we set out to understand the ways in which screen readers are problematic, particularly for novice users, and probe how we might design improved interfaces for non-visual access. We were particularly inspired by the way that early graphical interfaces appropriated physical metaphors with which users were already familiar and wondered if these same metaphors might be usefully re-appropriated into a nonvisual interactive system. We first conducted a field study of screen reader use at a school for people who are losing or have recently lost their sight. Based on the results of this work, we then developed a series of prototypes for multimodal nonvisual computing access, which we used in a short study with these same users. Use of physical artifacts enabled us to see new challenges in nonvisual computing not evident through our fieldwork.

This work offers two complimentary contributions to scholarly thinking about accessibility. First, we describe the challenges of nonvisual computing, with a particular emphasis on how we might use alternative input techniques beyond simply audio. Our in-depth exploration of this problem sheds light on underlying assumptions surrounding accessible software development and assistive technologies themselves. Second, we demonstrate the feasibility of an alternative multimodal approach and show the ways in which such systems might actually be easier and more efficient than screen readers, particularly for novice users.

The remainder of this article is structured as follows. We first overview the related literature in this area, both in terms of nonvisual computing and in terms of multimodal interaction more broadly. We then describe our approach to and the results of an empirical study of blind and low-vision users learning screen reading technologies. We then describe our approach to the design, fabrication, and evaluation of two exemplar multimodal interfaces for nonvisual computing. Finally, we close with design guidelines for multimodal nonvisual accessible computing and a discussion of the future of this area of research.

2 RELATED WORK

Significant efforts have been put towards improving the interaction space for blind computer users, but making sense of two-dimensional, spatially organized graphical interfaces (GUIs) using nothing but linear, ephemeral streams of descriptive text has proven difficult [12, 34]. GUIs simplify interaction for sighted users using physical and visual metaphors supporting direct manipulation (e.g., files and folders), but the digital objects that support these metaphors do not translate well to nonvisual use, making it dramatically more difficult than visual use [7]. Furthermore, GUIs support distributed cognition [17] by providing memory aids through information that can be directly perceived and by "anchoring and structuring cognitive behavior" [57].

By contrast, auditory speech interfaces rely mainly on text-to-speech processors that take text as input and return digitized speech as output. During the conversion process, visual information such as graphical emphasis and spatial arrangement is lost [4]. Work has been done to overcome this by augmenting the output with a navigation ontology [56] or increasing access to semantic information [31]. Additionally, speech input, non-speech input [19], and non-speech output (e.g., [11, 29, 40]) have been shown to enhance the interactive experience and the range of things that can be displayed. For example, sounds can be positioned in space, as with the hallway metaphor used by Schmandt [41].

The wide range of advances found in audio representations of screens can be accounted for in part by the ease of prototyping audio, which is primarily a software task. Even so, audio requires users to make a tradeoff between metaphors that are fairly closely tied to the hierarchy of GUI interfaces (e.g., [12]) or metaphors that differ so much that they require advanced expertise and learning (e.g., audio icons and contrasting voices [34]). Although it is commonly understood that visual impairment can lead to more efficient auditory perception [37], resolving accessibility issues through a single modality can still lead to unbalanced, complex engagements that are challenging for experts and non-experts. The large corpus of input commands required to operate audio interfaces impose significant hurdles as users first familiarize themselves, only to become a constraint on efficiency with increasing expertise [47].

One way to broaden the set of metaphors available is to move out of an audio-only realm and into the physical world of tactile interaction. The sense of touch is a rich input modality capable of perceiving a diverse range of feedback from size and shape, to texture, stiffness, and temperature [21]. Combining touch with computer interaction has been a focal point for researchers for much of the past two decades [44] and continues to hold promise as a bridge between the digital and physical world. The primary focus for nonvisual tangible output in assistive technology has been text display through the refreshable tactile output of braille displays. A lot of effort has gone into converting the rich set of visual desktop metaphors into something that can be controlled and displayed in the relatively limited world of a single 24- to 40-character braille display, mouse, keyboard, and speaker system (e.g., [29, 35]). However, the human perceptual system is much more sophisticated in its ability to remember and manipulate objects in the world. Our ability to remember the location of things in space, for example, is so good that this is often used as a memory aid for remembering a list [3]. Similarly, augmenting visual memorization tasks with tactile cues has been shown to be an effective memory aid for individuals with poor recall abilities [24]. For repeat tasks, spatial memory is further enhanced by proprioceptive capabilities, which are the ability to locate the relative position of body parts [13]. Thus, people have an ability to easily place things in space, remember their location, and remember repeat tasks (such as reaching repeatedly to the same location). In their seminal article, Ishii and Ullmer [20] explored how engagement with technology through physical spaces could lead to a "much richer multi-sensory experience of digital information."

The vision of Ishii and Ullmer [20] combined with an era of affordable digital fabrication has created an influx of research interest in the physical interaction space. As the devices they envisioned reached commercial viability, researchers discovered new ways of incorporating those devices into the physical world. For example, the metaDESK combines digital imagery with physical objects (phicons) using sensors and projection [20]. The emergence of tabletop multi-touch displays put the same functionality into a single device, allowing researchers to broaden the set of physical metaphors in new directions [52, 53]. Similar efforts have ranged from input only—such as the use of capacitive touch to improve mobile phone interaction [39]—to output only—using vibrotactile feedback to add texture to touch surfaces [5]. There have also been significant contributions towards developing novel forms of communication using the tactile channel. The AIREAL project successfully created the sense of movement around the body by using ultrasonic pulses of air [45]. Similarly, Gupta et al. [16] used air vortices to provide haptic response to gestural inputs. Using frequency modulation of an electrovibratory surface, Bau et al. [5] created TeslaTouch to demonstrate how an otherwise-smooth surface could produce a range of tactile sensations from stickiness and waxy to bumpy and rough.

More recently, tangible computing has gained interest as an assistive technology to augment the loss of the visual input channel. For example, TeslaTouch was repurposed as a tool for communicating two-dimensional (2D) tactile images to the blind [55]. Zuckerman and Gal-Oz [59] demonstrated that physically interacting with tangible objects elicits a sense of enjoyment—one that is preferred over traditional forms of interaction (e.g., keyboard and text-to-speech), even when the given task might take longer. This enjoyment is likely a result of the learnability of tangible interfaces due to their innate ability to be manipulated and explored [42]. However, matching the enjoyment of exploration to the complexities of a modern computing system remains a challenge. While multimodal sensory approaches that combine tactile and audio feedback have been shown to be effective at improving task completion times [23, 48], the amount of information that can be communicated is limited. Brewster and Brown [8] translated the properties defined for earcons [6] (frequency, duration, rhythm, and location) to the haptic channel in what they named "tactons." Tactons explored vibrotactile feedback as a source for generating abstract patterns to communicate complex messages nonvisually. Follow-up studies demonstrated that tacton patterns were distinguishable, though some patterns were more successful than others [9]. The TASO device made by Frank AudioData combined both the tactile and auditory approaches by using fixed vertical and horizontal sliders to move the cursor around a windowing computer system. The TASO used auditory cues to indicate when relevant content was reached using the sliders [32]. Prescher et al. [33] removed audio entirely by taking a literal approach to mapping graphical interfaces to the tactile modality with the BrailleDis9000, a 12-line-by-40-character braille display capable of rendering shapes and braille characters. Individual windows and widgets could be mimicked tactilely by raising specific pins on the device. Although the authors found the tactile representations difficult to grasp for some participants, most preferred the direct manipulation afforded by the device over input through keyboard commands.

The opportunity to explore computer interactions within the physical space is increasingly more approachable because of advances in low-cost rapid prototyping tools, such as 3D printing [22] and microelectronics [39], which can support fast and inexpensive production of the tangible objects and have already proven their value in the domain of assistive technology [18]. Thanks to these

Table 1. Students Observed for Field Study

ID	Gender	Age	Technology	Condition	Time with Condition
B1	Female	39	Screen Reader	Glaucoma	9 years
B2	Male	35	Screen Reader	Optic Neuropathy	10 years
B3	Female	43	Screen Reader	Retina Pigmentosa	since childhood
B4	Female	36	Screen Reader	Retina Pigmentosa	since birth
B5	Female	41	Screen Reader	Retina Pigmentosa	7 years
B6	Male	37	Screen Reader	Optic Neuropathy	4 years
B7	Female	41	Screen Reader	Retina Pigmentosa	since birth
M5	Female	34	Magnification	Glaucoma	8 years
B9	Female	45	Screen Reader	Glaucoma	4 years
I1	Male	52	Screen Reader	Unknown	since birth
M3	Male	35	Magnification	Cataracts	since birth

Detailed list of the visually impaired students and instructor observed at the field site. Technology represents the primary tool students preferred to use. Participant ID labels are coded according to technology; B group had less than 3 months of experience with their assigned technology, M group used magnification, and the I participant was the blind instructor.

advances, tangible interaction is increasingly accessible and appealing [44]. The blending of linear access, visual direct manipulation, search, and other dynamic mechanisms hold promise towards creating higher-quality experiences throughout all forms of interactive computing spaces.

3 FIELD STUDY OF NONVISUAL COMPUTING

We conducted a field study in a blind and low-vision computer class at a non-profit organization over a period of 4 months. This non-profit organization, EmpowerTech, provides computer-based job skills training (e.g., office productivity and Internet navigation skills) for people who are blind or losing their vision. The majority of students are relatively new users of assistive computing tools, such as screen readers and screen magnifiers.

Over the 4-month period of this study, the first author participated in 12 classes averaging 4 hours per class for a total of 48 hours of participant observation. Being both sighted and an expert user of screen readers and other accessibility technology, he participated in the classes as a teaching assistant, doing whatever tasks were assigned by the lead instructor, a blind man who is an expert computer and screen reader user.

Class size varied from week to week but typically consisted of 8 to 12 students (see Table 1 for student details). At the time the first author joined the class, experience with screen readers and magnifiers varied across students, but none of them had received more than 3 months of training. Field notes were taken during the class when possible and directly after. In class instruction focused heavily on screen reader operation and keyboard commands for common computing tasks, such as file management, web browsing, and word processing. Outside of class time, the students often stayed at EmpowerTech to practice their skills, providing additional time for observation of their computer use and for informal interviews. Field notes documented seating arrangements, lecture topics, software being used, and the interactions between students and the instructor. Recurring themes during lecture included the advantages and disadvantages of various assistive technologies as presented by the class instructor and the struggles and conveniences for the students.

The training classroom was rectangular in shape with two exits on either side. Tables lined the walls of the room and provided work areas for students to sit. Each table had two or three computers, mostly PCs running Microsoft Windows 7,[1] except for one table, which housed two

---

[1] https://www.microsoft.com/en-us/windows.

Apple iMac systems.[2] Sitting next to one of the iMac systems was a magnification screen that used an optics system to increase the viewable size of any object placed beneath it. In the center of the room there was a round table that was large enough to seat four students; however, it was often pushed up against one of the rectangular tables to keep the center of the classroom open for the instructor to move more freely as he taught. Many students brought their own laptops to class making the round table the preferred place to sit.

Classes at the field site are run Monday through Thursday from 9am to 1pm. The tools used within the classroom varied widely between students. Most students brought their personal laptops with them to class. During our time at EmpowerTech, there was only one student (B3) who did not have a laptop computer. Two students used Apple MacBooks (B5,M5), the remaining students used a mixture of Windows 7 systems. Additionally, two students transitioned to new systems: B6 migrated from Windows 7 to a MacBook, and B4 upgraded from Windows 7 to Windows 8.

Over the course of our 4-month study, the class progressed from learning how to use a screen reader to web navigation, email, and text editing. The first author typically collected data through classwide observation. On days when one or two students needed assistance beyond the scope of the lesson plan for the day, he would be asked to work directly with the struggling students.

Although the classroom had specialty devices available, like a digital magnifier and braille printer, neither were ever used. The primary tools that students relied on were the Victor Reader[3] and screen reader or screen magnifier. The instructor encouraged students to choose whichever operating system and assistive tools that they preferred. The Victor Reader, a handheld audio recording and playback device with text-to-speech capability, was used by all of the students. Students primarily used the device to capture verbal instructions from the teacher, which they played back at a later time while working on class assignments. However, the devices also supported audio output of music and electronic publications like PDF's and e-books. Although the core features of the Victor Reader could also be found on smartphones, the students that we spoke with found the tactile input through the raised buttons of the device faster and more convenient to use than a smartphone touchscreen. The most commonly used screen readers were JAWS[4] for students running Microsoft Windows and VoiceOver[5] for students on Apple MacBooks. A few of the desktop computers within the classroom were setup with multiple screen readers. In addition to VoiceOver, the iMac could also use the Dolphin screen reader. At least one of the Windows desktops had NVDA[6] installed.

Following each class, the first author spent time with individual students and conducted informal interviews about their experiences with technology both in and out of the classroom. Additional artifacts, including handouts and worksheets assigned by the instructor, were collected at each class period and saved for analysis. We paid particular attention to bottlenecks and challenges during instruction and work time. Between classes, the first author engaged with the most commonly used software in the previous class to better familiarize himself with the opportunities and challenges the tools provided and to prepare to support students during class.

Data analysis combined a mixture of inductive and deductive approaches. All field notes were read by the entire research team between each class, and discussions occurred about the notes during weekly meetings. Data were analyzed in an iterative fashion with a constant comparative approach. Each week, field notes focused both on the general themes and questions that prompted

---

[2] https://www.apple.com.

[3] http://www.humanware.com/microsite/stream/index.html.

[4] http://www.freedomscientific.com/Products/Blindness/JAWS.

[5] http://www.apple.com/accessibility/osx/voiceover/.

[6] http://www.nvaccess.org/.

our initial research and emergent questions from our analysis. We first examined our data using known challenges from the literature, such as the shortcomings of screen readers [7], software accessibility compliance [26, 36], and user frustration [25]. Using an iterative, inductive approach, we then identified additional emergent phenomena, named and categorized these issues. We used affinity diagramming and axial coding to understand the relationship between, across, and within these codes as well as to those themes from the literature and our original deductive coding. While completing our data collection and analysis, we iterated on the prototypes described in Section 5. As dominant codes emerged, they were incorporated into the categories presented in this article and considered as part of our prototype development process. This intertwined process enabled us to test the boundaries of technical feasibility as our ideas were broadened and then refined.

The first author also met regularly with the course instructor and school staff. These meetings supported three complementary goals of the research team. First, we were interested in building a strong community partnership with EmpowerTech, including the ability to hear from them what research questions they might have or technical innovations they might wish to see. Second, we conducted preliminary analysis and presented these initial ideas to them to inform the next data collection session. Finally, as the fieldwork drew to a close, we began to shift the discussion towards larger analytic themes and design guidance for our prototype solutions, again getting feedback from the members of the community we were studying and support from the organization with whom we were partnering.

4 RESULTS

In this section, we describe the results of our fieldwork. We first provide an overview of the motivation for students to participate in these courses to ground our understanding of the informants who worked with us to develop insights surrounding screen reader use and the potential for multimodal nonvisual computing. We then describe the challenges participants encountered in using screen readers, in terms of the dual issues of keystroke memorization (input) and audio interpretation (output). These challenges led us to an initial conceptualization of an alternative tangible based approach, which we describe in Section 5.

4.1 Drivers of Participation

Students in the training class we studied are enrolled primarily through the Department of Rehabilitation at the county level to meet a requirement for state-funded economic support. In many cases, these classes are simply required by state mandate. However, in other cases, students can be there for a variety of other, often more subtle, reasons that still nonetheless connect to their status as vulnerable and in need of government assistance in some way. One student in the classes we observed chose to be in the training program with no recommendation from a state agency. B6 had worked at his family-owned business for years prior to losing his vision. Unable to perform the functions he once did, he saw computers as a tool he could use to continue providing value to the business.

Given the high rates of government assistance, these training programs seek to help students to gain more independence. For example, major portions of the curriculum focus on practical skills like using government and commercial websites to find housing, employment, and home goods. When students have gained enough proficiency with their tools, they are assigned the task of using these sites to search for something of particular need or interest to them at that time. For example, while helping B2 work on one such task, the first author noted:

> *It's hard, but they do it because the services they rely on are increasingly web dependent. Although there are representatives available for students to directly communi-*

> *cate with, the use of the same web based services that sighted people use reduces the need for person to person communication.* —Field notes from the fourth week

This example brings to light the particular challenges that people with disabilities or low computer literacy—or both—are likely to have in an increasingly digital world with fewer resources in the nonprofit and government sector. Whereas a social worker or other government agent might previously have been readily available to work with a client seeking help, reduced funding and increased pushes for efficiency now mean that blind computer users are more heavily reliant on screen readers and computational technology.

Given that blind people are more likely to live in poverty than those with sight [2] and that the people in our class in particular had been sent by the Department of Rehabilitation, it is perhaps not surprising that most were unemployed and many lived with relatives. This context can make the push for autonomy and independence a challenging one. Likewise, the connection between the kind of tasks assigned in the class with the economic realities of the students was at times tenuous. For example, one day after 2 hours of frustration attempting to perform the assigned tasks using a screen reader, B9 pressed the instructor to allow her to use magnification software, prompting the instructor to declare:

> *"Magnification software can dramatically affect performance on your computers. Eventually you will need to rely on screen readers anyway, so you might as well learn it now." The instructor finished the lecture up with a short tour of how to navigate the Amazon website. B9 proclaimed, "I don't need to learn Amazon, I'm never gonna use Amazon. You have to have money to go shopping."* —Sixth week, participants I1 and B9

In this example, we see two important points of contention. First, the student's economic realities are being somewhat ignored with a "toy task" of shopping assigned. Given how challenging shopping online is when every single element that is returned must be navigated with keystrokes and audio, it is no wonder that students might find this task difficult. Add to it that they do not see any relevance of the task to their own lives, and you have a set of people unsurprisingly eager to try any other way. Second, for five students in our class, though they are predicted to lose their sight entirely in the near future, they still have some sight, which allows them to use magnification features on the screen rather than only the screen reader. What is notable here is that this particular student had to modify her text editor to prevent word wrapping so her magnification software could properly display the text she was reading, and yet this was still preferable to using a screen reader. In the face of this set of issues—a student who can still see, despises screen readers, and does not see the point of learning to shop using the screen reader—the instructor must push forward. He is not a masochist, far from it, but he recognizes that right now, with the current state of affairs, all people who will be blind simply must learn to use screen readers at some point. It is this context that we seek to change with this work.

4.2 Performance Constraints of Memorization

In lieu of the point-and-click interaction on which sighted computer users rely, nonvisual screen reader use is dependent on keyboard input commands. Encapsulating all of the actions required to perform visually oriented computing tasks for an audio-only interface increases the commands required to perform similar actions through direct manipulation. For example, in the fifth week, students were given a worksheet containing over 40 commands for operating a screen reader and Microsoft Word, ranging from document formatting to file management. Next to each task was a blank field in which students were required to enter the task that the command accomplished. The first author arrived that week shortly after the students had received these worksheets, and they were working through them, with some of them audibly complaining. Complaints ranged from concern that the commands would not be useful for those who still had some sight (as noted in Section 4.1) to the concern that the commands themselves were challenging to remember. At the end of the class, the first author discussed the students concerns over memorization:

> *When the instructor returned, I discussed what had happened with him. He acknowledged that there are a lot of commands to learn. "This is why everyone has a Victor Reader." The Victor Reader is a small recording and text-to-speech device that the students use to capture and playback the instructions.* —Field notes from the fourth week

This tension between learning what will be useful later and learning what is needed right now is a large issue in the teaching of assistive technologies [1, 38, 58]. The use of recording devices, such as the Victor Reader, points to the complexities of learning audio-only interfaces. The adage "recognition over recall," a fundamental tenant of human-computer interaction since the mid-1990s [30], is clearly lost when instructions for one system must be captured and replayed through another system. This issue is further confounded when navigating the variability across software systems. During one lecture the instructor responded to a question about the best web browser to use by saying:

> *"You need to learn all browsers because each one is better at a particular task. For example, FireFox is better at reading captcha's. I only use FireFox to pay bills, that is what it is good at."* —Seventh week, participant I1

The hurdle that memorization presents is not simply a result of being a novice user, the system should be given equal responsibility. Not surprisingly, faced with the option of using GUIs, even in a difficult and challenging way, students who were able to do so, always chose to use the visual interface over memorization. This choice demonstrates that the obstacle of learning to use the memorized commands may simply be too high for most novices, an issue we address directly with our tangible approach. One participant with low vision relied on her ability to perceive the motion of the focus indicator, a small dotted outline that surrounded the selected element, as she navigated around the various elements of a web page. Detecting the motion helped to orient her location within the page, after which she would rely on the audio output to select the desired element. Another participant was so adamantly opposed to using screen readers that he was willing to use a traditional glass magnifying lens coupled with a screen magnifier, a clumsy and painful solution, to avoid it. As disparate as these approaches seem, they are indicative of the students' preference for more than an audio-only interface.

4.3 Challenges of Audio-Only Output

Even with a reasonable amount of keyboard commands memorized, locating information on screen remained challenging for the students we observed. A common tactic employed by the students was to use the basic navigation controls like the tab and arrow keys to move between elements sequentially, a strategy identified by Vigo and Harper [46] as exhaustive scanning. With each element change, an auditory processing step is then required to hear and grasp the semantic description of the selected element and its associated information. For example, while attempting to download an email attachment in a web-based email client, a common difficulty among screen reader users [54], two students were using the tab and arrow keys to bring focus to the attachment link. Unknowingly, the command inputs that they were using prevented them from reaching the panel in which the attachment link was embedded. This resulted in a cyclical process of keyboard input and audio output that would restart after all the elements had been traversed. The students looped through the content numerous times searching for the attachment link before asking for help. Using these kind of strategies enables people to recover when they have forgotten a complex command sequence or to avoid learning such things in the first place. However, the recognition over recall strategy for primarily auditory output comes with substantial costs in terms of both time and frustration.

Additionally, as we see in the above example, the ephemerality of the audio stream may lead users to believe they have "missed" the link, when in fact they never encountered it. Laboring under this false belief, they take multiple passes linearly through the content. Eventually, users may recognize their inability to target the desired information, but even then, they may not grasp why. Without sight, recognizing that a particular piece of information is on another panel, which is relatively inaccessible to the screen reader in its current configuration, is nearly impossible.

Finally, the costs of this kind of traversal based seeking only grow with increased content. In the example above, each pass took an average of around 30s, with the users speeding up (from 60s on the first pass to 45s on the second pass, and so on) as they began to recognize repeated words. If we consider established average listening comprehension rates of roughly 190 words per minute [14], then traversal time for content-rich documents can quickly lead to unwanted delays. For example, in a similar situation observed during the sixth week of class, job searching in particular was challenging for participants as noted in our field notes:

> *B2 was attempting to search for jobs, with traversal of the results page of 20 job listings taking more than 10 minutes to complete. With a much larger content space the traversal time would increase to unusable proportions. In this particular case, the traversal time between the beginning and end of the results was long enough that B2 was not able to recognize that the traversal had looped back to the beginning.* —Seventh week, participant B2

Not understanding content scale and excessive audio processing posed formidable constraints to efficiently accomplishing basic computer tasks.

The time cost of seeking is further confounded by the complexity of the user interface, where *complexity* represents an increase in non-linear text arrangement. During another conversation with M3 on his preference for magnification software, his dislike of the screen reader was further explained:

> *"That's another reason I don't like the screen reader, is it sometimes reads stuff I don't even need it to read."* —twelfth week, participant M3

Where non-linear arrangement benefits sighted users by placing tertiary content to the side, the screen reader treats all text equally, sending undesired information to the user for processing. For example, a table provides a convenient, easy-to-process visualization of categorical data. However, when targeted by a screen reader, a detailed description of the structure and arrangement must be communicated before any of the actual data. The following quote is the text-to-speech output of a web-based email inbox using the NVDA screen reader:

> *"Table with thirty-six rows and four columns, row one column one, checkbox not checked, column two link mom."*

In this example, 18 words must be spoken before the user can identify that the first email in the inbox is from "mom." Although shortcuts to navigate around the inbox are typically available, many of them are unique to the application being used and difficult to master for all but the most advanced users. Still, the benefits of committing these immense command sets to memory is only as beneficial as the software infrastructure allows. One of the more advanced students at EmpowerTech (M5) was forced to use the seeking strategy to traverse results from a housing search when the website she was using failed to properly use heading tags to organize the search results. A simple mistake in formatting made it impossible to use the screen reader shortcuts she had learned. Here, knowing that there was an easier way, but that it failed to work, was incredibly frustrating for this student. Her solution was to print the listings in a very large font and rely on her limited vision or the help of a friend to read the results. From her perspective, this was preferable to having to repetitively seek the information.

A wide variety of challenges exist beyond simply searching through a dense, ephemeral, and linear audio stream. In particular, input is intrinsically also an act of output for sighted users, though we rarely recognize it as such. As we type words into the screen, we see them appear letter by letter. Anyone who has ever been dogged by slow response time from a website, piece of software, or operating system can attest to the challenges of typing when whole words or sentences show up rather than character by character. Nonvisual users, however, do not have this benefit available to them. For example, in one instance, P2 struggled to understand why a job site was not able to find any jobs using his queries. The issue was easily identifiable to a sighted observer: The search field had identical search terms concatenated together. P2 was performing the task he had been trained to do, but when a shortcut designed to reduce repetition for visual users (prepopulating a text box with last used text) failed to notify P2, his workflow fell apart. These kinds of shortcuts do indeed make interaction faster and more efficient for sighted users. However, they are not visible to those using screen readers. Thus, it became clear to us during this fieldwork that two-way communication with the computer would be required rather than the parallel but unidirectional flow of keyboard-only input and auditory-only output.

The context of the computing interaction can further exacerbate the issues of audio output. Our fieldwork was particularly helpful in understanding how screen readers might work in an open work environment, the kind common to many office spaces, particularly in the United States. The training program primarily took place in one small classroom as described above. Although each student had an individual workspace, and no computers were shared except when students were explicitly collaborating, the noise in the room still rose to distracting levels during every observed session. Distractions were common as individual students entered and left the room or talked with each other. Similarly, the instructor needed to listen to the students' screen readers to make sure they were pacing appropriately with the material as well as to provide assistance when difficulties arose.

What we see from this experience is a larger issue related to distractions in general. As noted in prior work [10, 15, 27], distractions in the workplace can be problematic for any computing user. Following a distraction, individuals must reorient to the original task, an activity that requires a certain amount of cognitive load and usually involves glancing across the workspace. In the case of nonvisual computing, however, we observed additional challenges to the reorientation task following a distraction. Not able to quickly glance through the screen, participants had to relisten to some amount of audio. The amount required can differ, leading to users sometimes listening to a short portion of audio, backing up further and listening to a longer portion (inclusive of the short portion they just heard), and so on, sometimes doing this several times before they can find their place. At times, this can be even more problematic, if the distraction led them to make an error, thereby navigating them to a different page or even a different application without their knowledge. Although sighted users bump their mouse or make errant keystrokes regularly when distracted, these are relatively easy to overcome using visual inspection. Without that option, however, we see even greater challenges. These results reinforce the need for a way to easily mark one's place when interrupted and avoid relistening to substantial portions of audio.

4.4 Discussion

Once we recognize the difficulties that audio can introduce to a nonvisual computing system, we can begin to identify alternative modes of interaction. Realistically, a nonvisual system will always be dependent on audio output in some form; however, by reducing the amount of audio required, we might be able to improve the computing experience for blind users. Our fieldwork demonstrated that memorization, ephemerality, and repetition produced negative effects that led to reduced computing performance and enjoyment.

As we described in the previous section, locating the desired information using a screen reader can be difficult, due in large part to loss of the ability to directly manipulate objects on a screen. For example, given a collection of icons on the desktop, a sighted user would identify the desired application by its shape and color, both artifacts that help it stand out. When accessed with a screen reader, the collection of icons is converted to a list of words that must be communicated linearly without any special identifiers. In the next section, we look at how these visual identifiers can be re-introduced through additional modalities.

5 THE OPPORTUNITY FOR TACTILE AND MULTI-MODAL INTERACTIONS

Tactile interaction makes it possible to replace the ephemeral nature of the audio stream with a permanence similar to that of a graphical display. Building on our earlier example of icons, if an icon were placed in the physical world and bound to the system state such that when an icon corresponding to an application is open, closed, or active, its physical state could change to match. In this way, inferring meaning through shape and location, the condition of the system state could be accessed directly by the user through touch without processing an audio stream.

We set out to explore this space by creating peripheral devices that target two common desktop computer activities required for almost all computer use: switching and locating. Our "Tangible Desktop[7]" system is composed of physical implementations of the computer taskbar and application window scrollbar (see Figure 1) that were built with inexpensive rapid prototyping tools. In this section, we describe the potential design space for multimodal interactions as well as our prototype system developed to test out that design space.

5.1 Design and Function

The peripherals we created for the Tangible Desktop were built around a motorized slide potentiometer, a device commonly found in audio mixing boards, using low-cost electronics and a 3D printer. The physical interaction of slide potentiometer enforces directional motion along a single axis, similar to the TASO device made by Frank AudioData [32]. Unlike the TASO, however, our devices are not designed to move the mouse cursor but rather traverse heirarchical content structures. Furthermore, the combination of a potentiometer for data input and a motor for output in a single prefabricated unit allowed us to explore the benefits of maintaining a synchronized state between the computing system and the user's cognitive model through bidirectional control. The potentiometers were controlled through a single Atmel 328–based microcontroller that relayed commands over USB to a host computer running Microsoft Windows 7.

The Tangible Taskbar uses the potentiometer to switch between different computing entities (i.e., files, programs, and browser tabs). Entities are represented by physical icons, conceptually inspired by metaDESK's phicons [20]. However, in place of the optical and electromagentic sensors used in metaDesk, our icons are implanted with radio frequency identification (RFID) chips (see Figure 2). The physical icons were 3D printed with a 2mm-deep crown that was filled with moldable rubber. The rubber crown for each icon was given a unique pattern of indents and

---

[7] Visit http://markbaldw.in/tangibledesktop/ to see a short operational video of the Tangible Desktop.

Fig. 1. A picture of the Tangible Desktop in its standard arrangement. The Tangible Taskbar sits to the left of the laptop while a user engages with the thumb of the Tangible Scrollbar.

Fig. 2. The physical icons used in the Tangible Desktop (see Figure 1). Each icon has an RFID tag embedded inside and a tactilely distinct rubber crown.

ridges. The surfaces allowed for tactile differentiation between icons. An inexpensive, commonly available RFID reader is mounted to the slide arm of the potentiometer, allowing the RFID chips in each icon to be read as the arm is moved. A 3D-printed housing is used to enclose the hardware and provide a slotted tray for placing the physical icons. Communication software that runs on the host computer manages the binding between RFID chip identifier and the desired computer

Table 2. Semantic Processing Overhead

Screen Reader	Screen Reader with Haptic
"Heading level one, Introduction"	"Introduction" + tap
"Link next page"	"Next page" + vibration

A comparison of screen reader speech output between auditory and haptic semantic descriptions. Note that haptic responses can occur in parallel with speech output.

entity. A physical button on the Tangible Taskbar sends a binding command to the host computer, instructing the communication software to assign the active window to the RFID chip's unique identification number. Once a binding is created, the software transitions the bound entity to an active state whenever the identifier is read by the RFID reader. For example, if a physical icon is bound to a browser tab containing the University of California homepage, when the slide arm is moved inline with that physical icon that tab will be given focus. If the desired entity is already in focus, then the environment will not change. Likewise, if a different bound entity is given focus through external means (i.e., mouse or keyboard), the motor attached to the potentiometer will move the slide arm to the corresponding physical icon.

The Tangible Scrollbar uses the potentiometer to traverse content in an application. A small eccentric rotating mass vibration motor is mounted to the slide arm to provide vibrotactile feedback to the hand. While other projects have explored the richness of vibrotactile feedback (e.g., [9, 23, 55]), our system intentionally limited the amount of information communicated through vibration to avoid many of the complications that accompany vibratory patterns. Tactile push buttons are mounted at either end of the device to provide additional input capabilities while using the scrollbar. Like the Tangible Taskbar, a 3D printed housing encloses the hardware and provides a graspable bar for pressing the buttons from multiple hand positions.

For our preliminary exploration, we limited the scope of the Tangible Desktop to switching tabs and traversing content in a web browser. We leveraged the extensibility of the Microsoft WebBrowser Control[8] and Speech API[9] to create a custom screen reader application. This approach allowed us to generate a testable environment without having to build fully functional drivers and software.

5.2 Discoverable Information

Information that is output from a screen reader is paired with a semantic description of the lexical information that has been requested by the user. This structure places two levels of cognitive processing on the user. First, they must comprehend the type of information; then they must process that information. For example, a document may contain multiple heirarchical levels, varying content lengths, and hyperlinks to other documents. Using a screen reader, the headings (or hyperlinks) are specified semantically (e.g., "heading level 1 Introduction"), and content length must be requested using input commands. By moving this semantic information out of the audio channel and into the tactile realm we reduce the amount of audio the user must process. Furthermore, this allows us to communicate the semantic and lexical in parallel, reducing the overall processing time (see Table 2).

The Tangible Taskbar is a physical representation of the desktop computer taskbar that places the basic operations surrounding a single application into a real-world, tangible object. We built

---

[8] https://msdn.microsoft.com/en-us/library/aa752040%28v=vs.85%29.aspx.

[9] https://www.microsoft.com/cognitive-services/en-us/speech-api.

our prototypes using 3D printing, moldable rubber, and RFID tags to create tactilely identifiable icons. With the support of an RFID reader, the icons can be individually bound to a single entity within the computing environment.

The taskbar provides a space to rest individual RFID-based icons, which can be bound to applications on the host computer. As an icon is placed on the taskbar, the associated application is launched. When the icon is removed, the application is closed. A slider located behind the taskbar indicates which icon has focus. A user can switch between applications by moving the slider into alignment with the desired icon. If the active window is changed through a different input (e.g., keyboard or mouse), then the slider automatically moves into alignment with the icon bound to the newly active window.

5.3 Data Permanence

The Tangible Scrollbar is a navigational device similar to the scroll wheel found on most computer mice. It is used to linearly traverse and navigate content while simultaneously communicating visual and semantic information tactilely, thereby supporting the kind of two-way communication and permanence required to help users find their place in documents and remember their last place. Unlike the mouse scroll wheel, which uses a free spinning motion, the Tangible Scrollbar has distinct endpoints that physically indicate the beginning and end of scrollable content. As content is traversed, the scroll handle delivers resistance and vibration to communicate semantic information to the hand.

To accommodate varying lengths of content, we calculate a distance of travel ratio that is spread equally between the scroll endpoints. This ratio is then used to determine how far the scroll thumb needs to travel before providing feedback. To reduce dependency on a mouse or touchpad, the Tangible Scrollbar also has two buttons that can be programmatically assigned to perform the same tasks as a mouse button.

When a web page is loaded for the first time in the custom browser, a reset signal is sent to the scrollbar, prompting it to move the scroll thumb to the home position. As the scroll thumb is moved, the web browser traverses all HTML elements that contain lexical content. The lexical data are delivered through the screen reader as speech, but the semantic data iare passed to the Tangible Scrollbar and delivered haptically. When a navigable element is encountered, the user can interact with it by pressing the navigation button. If this interaction updates the content (e.g., navigating to a new web page), then the scroll thumb automatically repositions itself to the top of the page. Likewise, if the previous content is restored, the scroll thumb automatically repositions itself at the last known location.

In preliminary tests, the frequency of tactile feedback had the side effect of also communicating content size. Content size is typically delivered to the user visually by changing the height of the scroll thumb of the graphical content pane. Similarly, our Tangible Scrollbar would reduce the distance of travel between haptic feedback events as the content size of the page increased, thereby providing the user with a sense of size.

6 EXPERIMENTAL VALIDATION

We conducted an experimental study coupled with qualitative interviews to understand how the Tangible Desktop compares to traditional computer interaction for sighted and visually impaired users. We recruited 16 participants (5 sighted, 8 low-vision, and 3 blind) through word of mouth (S1, S2, S3, S4, S5), university services (M1, M2), and our field site, EmpowerTech. The study was conducted in a usability lab at the University of California, Irvine (n=5 all sighted users, n=2 low-vision but legally blind users) as well as the field site in Los Angeles (n=3 completely blind users, n=6 low-vision but legally blind users). Table 3 provides a detailed view of the participants,

Table 3. Study Participants

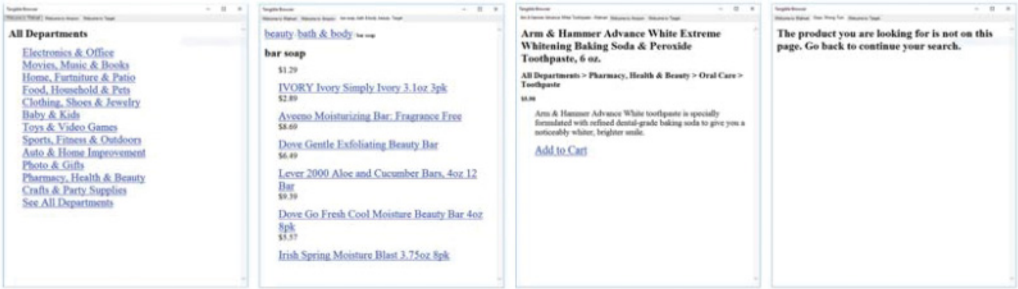
ID	Gender	Age	Technology	Braille Reader	Field Study
B1	Female	39	Screen Reader	–	√
B2	Male	35	Screen Reader	–	√
B3	Female	43	Screen Reader	–	√
B4	Female	36	Screen Reader	–	√
B5	Female	41	Screen Reader	–	√
B6	Male	37	Screen Reader	–	√
B7	Female	41	Screen Reader	–	√
I1	Male	52	Screen Reader	√	√
M1	Male	18	Magnification	–	–
M2	Male	20	Magnification	–	–
M3	Male	35	Magnification	–	√
S1	Female	27	Display	–	–
S2	Female	28	Display	–	–
S3	Male	37	Display	–	–
S4	Female	37	Display	–	–
S5	Female	26	Display	–	–

Technology represents the system participants typically use. Participant ID labels are coded according to technology; B group had less than 6 months of experience with their assigned technology, M group used magnification, S group were the sighted participants, and the I participant was a blind assistive technology instructor.

including the crossover between our study and field work. Eight participants identify as women with an average age of 35 (SD=8.79). Three participants (M1, M2, and M3) use magnification on their personal computer systems.

6.1 Procedure

Our fieldwork demonstrated that browsing the Internet is a critical stumbling block for people with visual impairments. Thus, our study used Internet browsing performance as the primary outcome measure for understanding the potential efficiency and experience of using a traditional screen reader as opposed to our prototype multimodal system. The current capabilities of our experimental system are limited to simplified HTML parsing, which made using live websites unreliable. We created three websites with navigational hierarchies modeled on the popular Internet shopping sites Walmart, Target, and Amazon. These sites were selected for their brand recognition as well as their incorporation into classroom assignments throughout our fieldwork. The sites that we built were modified to only include the navigational structure, an error page, a product list page, and product detail page (Figure 3). Each product list page contained five comparable products in addition to the target product for the task. However, because the site navigation was modeled after the commercial sites, the number of navigable elements accessible by the Tangible Scrollbar varied among target products and each website (see Table 4). All additional website content such as product recommendations, customer reviews, and advertisements were removed. The error page was displayed when a participant selected a product category link that did not lead to one of the task products. The two products that we used were selected from common household items (toothpaste and bar soap) available on all three Internet shopping sites. A product page was created for each target product that reflected the headings, description, and price used by the actual shopping site that sells the product. In instances in which the actual product prices were identical, we modified the price to ensure only one of our study sites had the lowest price. All the web pages that we



All Departments

Electronics & Office
Movies, Music & Books
Home, Furtniture & Patio
Food, Household & Pets
Clothing, Shoes & Jewelry
Baby & Kids
Toys & Video Games
Sports, Fitness & Outdoors
Auto & Home Improvement
Photo & Gifts
Pharmacy, Health & Beauty
Crafts & Party Supplies
See All Departments

beauty bath & body bar soap

bar soap

$1.29
IVORY Ivory Simply Ivory 3.1oz 3pk
$2.89
Aveeno Moisturizing Bar: Fragrance Free
$8.69
Dove Gentle Exfoliating Beauty Bar
$6.49
Lever 2000 Aloe and Cucumber Bars, 4oz 12 Bar
$9.39
Dove Go Fresh Cool Moisture Beauty Bar 4oz 8pk
$5.57
Irish Spring Moisture Blast 3.75oz 8pk

Arm & Hammer Advance White Extreme Whitening Baking Soda & Peroxide Toothpaste, 6 oz.

All Departments > Pharmacy, Health & Beauty > Oral Care > Toothpaste

Arm & Hammer Advance White toothpaste is specially formulated with refined dental-grade baking soda to give you a noticeably whiter, brighter smile.

Add to Cart

The product you are looking for is not on this page. Go back to continue your search.

Fig. 3. Sample of the study websites that were used, from left to right: top-level page for the Walmart site, product listing page, product detail page, and error page.

Table 4. Website Elements

	Walmart		Amazon		Target	
	Toothpaste	Bar Soap	Toothpaste	Bar Soap	Toothpaste	Bar Soap
Home Page	14	14	42	42	13	13
Level 1 Page	35	35	9	9	32	32
Level 2 Page	14	10	39	39	12	9
Level 3 Page	13	14	13	13	16	16
Product Page	5	5	5	5	6	3

A list of the total HTML elements processed by the Tangible Desktop for each website and product used in the experimental study.

created were constructed using accessibility standards including screen reader-dependent HTML tag attributes.

Each participant completed two shopping tasks, one for each product, using the same three websites. One shopping task was completed with the experimental Tangible Desktop, and one task was completed using the participants personal system. All but one participant used their laptop computer to complete the task. Participant B3 did not own a laptop, so she used a familiar desktop computer in the classroom. Participants were allowed to use any assistive devices that they normally used in conjunction with their personal system. The starting system and the product to shop for were randomly assigned (see Figure 5). When participants performed the task using their computer they were asked to use their current typical web browsing environment.

We did not conduct a formal training session prior to the start of study. Participants were given a minute to locate the experimental system and explore it tactilely. Tactile exploration occurred naturally; we did not explicitly instruct them to do so. The PI then explained that they would move the thumbs on each device to switch between and navigate across the websites. The PI also explained the meaning of the haptic feedback indicators that the Tangible Scrollbar provided. Before beginning the task, the address of each shopping site would be spoken out loud to the participant so they could preload each website. Once all three sites were loaded, the assigned product name would be spoken verbally. Participants would then be asked to find the lowest price for the product across the three shopping sites. The task was considered complete when the product with the lowest price was added to the shopping cart for the site.

When completing the task using our experimental system, navigation and interaction was performed with the Tangible Scrollbar while switching between the shopping sites was performed with the Tangible Taskbar (see Figure 4). The three shopping sites were each bound to their own physical icon by the primary investigator, prior to the start of the study. Auditory feedback was

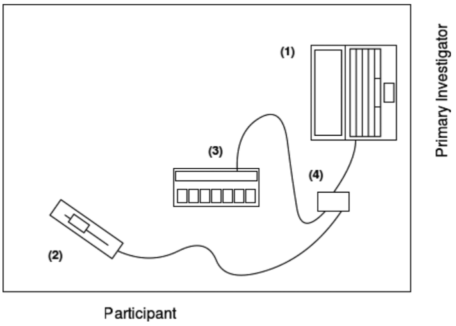


Fig. 4. The study setup: (1) PI system running study software, (2) Tangible Scrollbar, (3) Tangible Taskbar, and (4) Controller interface.

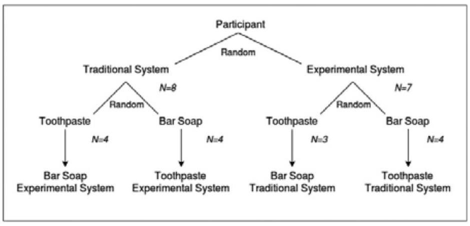


Fig. 5. Flow diagram showing the experimental design. The starting system was randomized at the beginning of the study. Once the starting system was selected, the starting task was randomized. The study was completed using the remaining system and task. Participant B7 is excluded due to personal time constraints.

Table 5. Screen Reader Configuration

HTML Element	Screen Reader Output	Haptic Feedback
<H1>...<H4>	Heading Text	Tap
<P>	Paragraph Text	Short Tap
<A>	Link Text	Vibration

The HTML elements controlled by the Tangible Desktop for the pilot study and their associated system responses. Screen Reader Output describes the type of content that was conveyed through computer audio using text-to-speech. Haptic Feedback describes the vibrotactile sensation delivered by the Tangible Scrollbar.

delivered through a custom screen reader designed to only render speech for lexical information on a subset of HTML elements. Table 5 lists the elements that we controlled, the text that was output, and the associated haptic feedback rendered by Tangible Scrollbar. As each website was explored, the Tangible Scrollbar automatically reoriented itself to the last known location of a given page. When an unvisited page was selected, the scroll thumb would move to the top of the device. On visited pages, the scroll thumb would move to the last known location prior to leaving the page. On completion of the study, participants were given an opportunity to describe their experience using the Tangible Desktop.

6.2 Analysis

We used time-based task performance data and post-study interview data for our analysis. We measured overall performance by comparing the task completion times between each participant's normal computing system and their use of the Tangible Desktop. Sighted participants were averaged together and served as a baseline to isolate the task length and complexity. Low-vision participants were excluded from our statistical analysis due to their preference for screen magnification. We also excluded the instructor from our statistical analysis. He was a statistical outlier probably because he had been using a screen reader for more than 20 years and had a vast experience to draw from when navigating our interface. Thus, participants B1–B6 were included in our statistical analysis.

The start and end of each study sub-session was extracted from the video log. Reviewing the video also allowed us to remove time corresponding to interruptions or errors that occurred. All but one participant experienced interruptions due to experimental system crash, phone calls, dropped Internet access, and questions about the task. In these situations, the first author used video

Table 6. Study System Error Corrections

Participant	Time Correction (s)	Reason
B1	145	Study system bug.
B1	12	Participant stopped to tell story.
B2	340	Study system crashed.
B4	96	Study system crashed.
B5	332	Internet connection at study site dropped.
B6	35	Participant stopped to make comments on product pricing.

A list of error corrections, including number of seconds that were removed from the final task time and the reason for the delay, that were made for affected participants.

recording and screen recordings taken during the study to filter the delays from performance data (see Table 6). A delay time in milliseconds was captured by taking timestamps at the beginning and end of each interruption in the video recording. The delay time was then subtracted from the time in the data logs at the event at which the interruption began.

The group used for statistical analysis consisted of six blind or low-vision participants, recruited from the training class described in our fieldwork, and thereby primarily novice screen reader users. Two students did not complete the entire study (B1, B7). While B1 was able to complete the task using the experimental system, she was unable to complete it using her system. We have included B1's data in the results but capped the participant system time at 20 minutes. By capping the participant system at 20 minutes, we actually reduce the likelihood of finding statistically significant improvement with our experimental system, thereby limiting the bias that such a choice might create. B7 was unable to complete the experimental system task due to personal time constraints. Thus, we used only her qualitative feedback in our analysis.

The experimental setup of our study introduced some limitations. Notably, our results indicate that the Tangible Scrollbar became harder to use accurately as the number of tactile interactions increased. The physical endpoints of the slide mechanism kept the distance traveled from top to bottom constant, but the distance between interactions decreased as the number of page interactions increased. During pre-study tests, we determined roughly 40–50 interactions to be the maximum that the device could traverse while maintaining a reasonable resolution. However, we did not take into account the various gripping strategies that our participants engaged while using the Tangible Scrollbar. When participants used a lighter grip, the haptic feedback mechanisms were too strong and moved the thumb into an unintended position. Similarly, the custom screen reader we built for the experiment did not provide a mechanism for increasing speech rate in the text-to-speech engine, a common modifier used by screen reader users to increase text processing performance. This suggests that we may be able to observe greater performance gains by adding support for increasing rate of speech.

Our qualitative results were collected from post study interviews. Results reflect the experiences of all participants except for B7, whose time constraints prevented us from conducting an interview. We asked each participant to describe their experience using the Tangible Desktop. Additional probing questions were asked in response to the answers that we received. Our follow up questions varied between participants but were patterned to elicit information on the strengths and weaknesses of our system.

6.3 Results

Our results indicate that screen reader users can work significantly more quickly using a multimodal system over traditional screen readers. Additionally, this improvement can be seen within a very short time of using the system, with limited training. Comments from participants suggest

Table 7. Participant Completion Times per Trial

Participant	Participant System	Experimental System	Starting System
B1	20:00	09:37	Experimental
B2	12:58	09:55	Experimental
B3	14:56	08:26	Participant
B4	16:32	12:37	Experimental
B5	19:54	09:41	Experimental
B6	14:43	10:03	Participant
Average	**16:29 (16.49 mins.)**	**10:03 (10.05 mins.)**	

Task completion times in minutes for the six participants included in the statistical analysis. The starting system indicates which system the participants used to complete the first task (randomized).

Table 8. System Performance as a function of System Type

Variable	Coefficient (seconds)	Standard Error
System type (1=experimental system, 0=participant system)	**-367.90^{**}**	**(84.19)**
Website order (1=Target first)	-116.60	(152.30)
Intercept	990.50^{***}	(56.76)
Adjusted $R^2 = 0.67$, $N = 6$	* p < .05; ** p < .01; *** p < .001	

Results of linear regression model examining the effects of system used on study completion time, controlling for website order (system order was randomized).

that greater personalization and customization as well as additional experience with the devices would improve the experience. In this section, we describe the results of both our statistical and qualitative analyses, highlighting the ways in which multimodal systems appear to improve user experience for nonvisual computer users.

On average, screen reader participants completed their tasks in 10.05 minutes with the experimental system as compared to 16.49 minutes using their own systems, an improvement of 39.0% (*(16.49 - 10.05) / 16.49 = .390*), $t(5) = 4.94$, $p < 0.01$. Table 7 details the time spent, in minutes, completing the task on both systems. Task completion time improved for *every* participant (see Figure 6 and Table 7). We randomized which system participants used first. Even though four of six participants used the experimental system first (see Table 7), their performance was still faster on the experimental system despite having no prior experience performing the task. Because we did not randomize which website participants used first, we controlled for this factor in our regression model. We found no significant effect for website order (see Table 8). Even after controlling for this potential confound, we still found that using the experimental system was significantly faster for screen reader participants.

To test the breadth of such improvement, we also looked separately at results for sighted participants and those with minimal sight who use magnification to augment their screen reader use (see Figure 7). Sighted participants were able to complete the task faster using their traditional system. This was, as expected, due to their preexisting familiarity with visual navigation of web pages. We saw a similar but smaller trend for magnification participants.

Not only did the quantitative analysis of our tangible system improve computing performance when compared to the audio-only interface of the screen reader, many participants described the Tangible Desktop as easier to use and understand. The 10 participants who were visually impaired commented positively about the use of the physical devices. For example, one blind participant expressed excitement on completing the experimental system task:

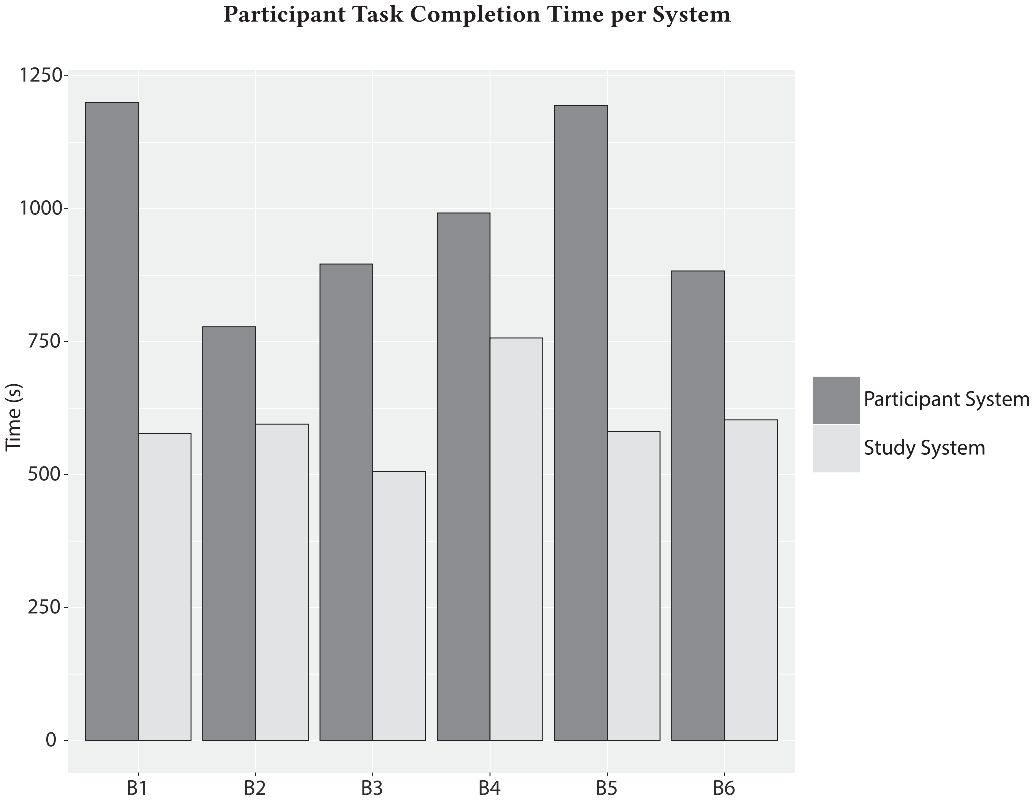


Fig. 6. Task completion time comparison between the participant system and experimental system for each participant. Completion times are total elapsed time, so we have not included error bars.

> *"I think it is easier, because blind people, they do things by touch. So they are very sensitive by touch. So if they touch things they remember faster. For me I would remember faster."* —Participant B3

Even the participants who preferred magnification acknowledged the benefits afforded by a tactile interface:

> *"Definitely more tactile than a software program, that was nice. There was enough pressure that I could tell what I was doing."* —Participant M2

There was an initial hesitation among all the participants as they familiarized themselves with the Tangible Scrollbar. The lack of pre-study training for the experimental system left several participants feeling like they could perform better if they performed the task again. Some participants explained at the end of the study,

> *"In hindsight, if I was going to go through that again I would be better the second time around."* —Participant B6

The novelty of the device, from the automated movement of the scroll thumb to the haptic feedback that it provides, was a likely cause for this hesitation. Yet the brief period of orientation in comparison to the weeks and months of screen reader training the participants have received

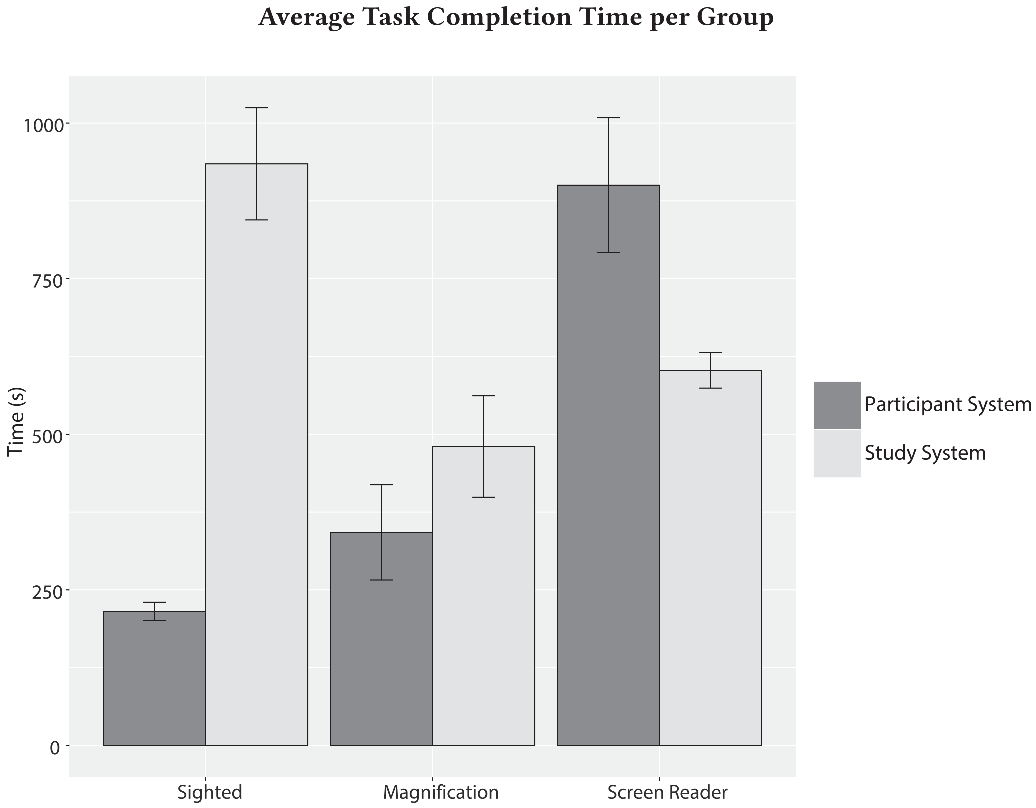


Fig. 7. Average task completion time comparison between the participant system and experimental system grouped by technology.

would seem to indicate that a tactile experience is less challenging to learn. For example, the first time participants performed a navigation event, the scroll thumb's automated reorientation caught them off guard. For most, their fingers were either still on or near the thumb when it moved, preventing it from completing its programmed movement. However, after a few navigation events, participants settled into a pattern of "click and hover," relaxing their hand slightly above the thumb until it had reached its destination.

Although all participants appeared to learn quickly and to enjoy using the Tangible Desktop, some did experience confusion over the functionality of the navigation buttons on the Tangible Scrollbar:

> *"Like I don't know what the dot is or the double dot, but if I use it a couple of times, I would remember."* —Participant B3

We did not specify how to hold the Tangible Scrollbar at any point during the study. The grip style selected by each participant varied widely and was never used as we imagined. Our preliminary design for the Tangible Scrollbar assumed that it would be held in the palm of one hand and controlled with the forefinger and thumb of the opposite hand. Instead, the device was left in place on the desk as it was positioned prior to the start of the experimental system task (see Figure 8). This created some initial confusion over the functionality of the Tangible Scrollbar. The symmetrical shape of the device initially left participants confused over vertical movement direction (up

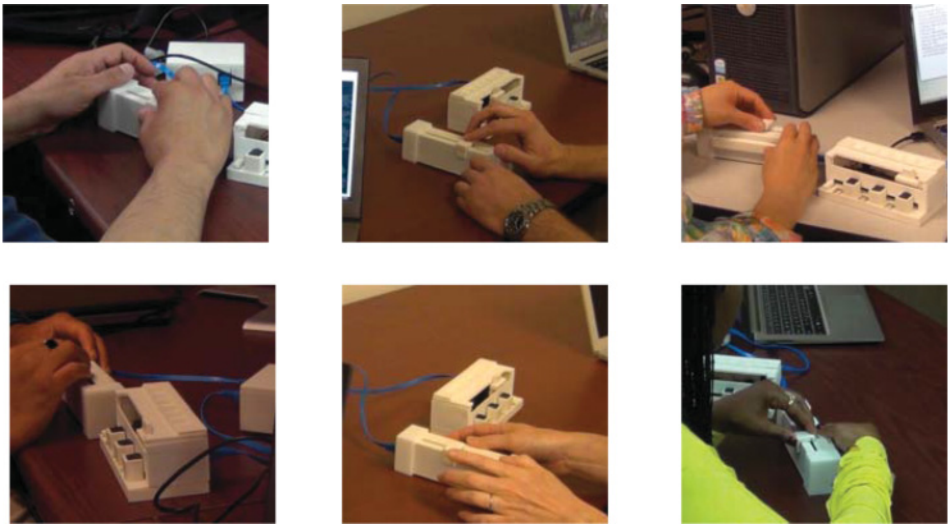


Fig. 8. Examples of the different hand positions used by study participants.

vs. down) and forward and backward button navigation. Therefore, an orientation period typically occurred during the first few seconds of use.

During post-study data analysis, we observed differences in how seeking was approached between the Tangible Desktop and the participant system among the novice group. While participants generally proceeded with caution while using their system, the experimental system was used without hesitation once the initial functionality of the system was understood.

7 DISCUSSION

The results of our experiment suggest that introducing additional modalities for interaction to blind computer users holds promise for improving enjoyment and efficiency for general computing tasks. While the small study population and sample diversity were not broad enough to generalize our results across all blind computer users, they do indicate that a multimodal approach to nonvisual computing is an effective compliment to audio-only output among novice users.

As we described in our review of related work, the research field has explored multimodal and tangible forms of computer interaction in great depth. Individually, concepts such as positional constraints [32], haptics [9, 55], and physical icons [20] have been shown to be meaningful for nonvisual and sighted users alike. Through the Tangible Desktop, we have arranged many of these concepts in a platform that lets us examine new forms of engagement with digital information. For example, the TASO device employed a similar strategy to our Tangible Taskbar and Tangible Scrollbar by using vertical and horizontal sliders [32]. Our system diverged from the TASO by purposefully applying the lateral motion that the slider provides to specific contexts. These contextual constraints (e.g., the Tangible Scrollbar only traversing content) are one likely reason participants approached our system with less hesitation than their own when navigating through content. Unlike the keyboard, making a mistake with the Tangible Scrollbar does not result in unexpected changes system wide. Like the TASO, the BrailleDis9000 approached nonvisual interaction through literal adaptation; translating graphical interactors such as windows, buttons, and scrollbars into tactile representations [33]. The Tangible Desktop takes a similar approach, but instead of creating a literal *one-to-one* representation (i.e., representing a window with a square) of graphical elements, we created *conceptual* representations capable of accomplishing the same task. The combination of the conceptual and contextualization made the Tangible Desktop less reliant on vibrotactile feedback for non-auditory communication. Managing a smaller set of distinct patterns in turn eliminated the need for long training sessions, avoiding many of the pitfalls that Brown et al. [9] found with Tactons [9]. Realistically, our vibrotactile pattern set (see Table 5) is far too small to be usable in a fully functional environment; however, our multimodal approach provides enough flexibility to manage the balance among taction, vibration, and audition.

Although the results of our study are positive, the Tangible Scrollbar system introduces numerous complex problems that need to be solved. The middleware that was developed is only capable of processing basic HTML elements and attributes, preventing the current implementation of our browser from functioning properly on modern websites. Larger documents with identifiers numbering in excess of 40 increase the resolution of the scroll thumb so much that it becomes difficult to move without skipping over items in the list. This could be resolved by organizing content into a paging system that splits identifiers into groups when they exceed the maximum supported number. Keyboard arrow keys could then be used to move between page groups. Alternatively, a user could toggle between different types of identifiers (e.g., links, form inputs, headers, and paragraphs) to narrow the scope of the scroll thumb to one group at a time.

One concern with this approach is the risk of increasing the demand on cognitive load by moving recognition from the visual channel to the tactile. Through our study with sighted users, we observed some difficulty in recalling the mapping created between tactile identifier and its associated semantic meaning. When scaled to an everyday working environment with a larger set of identifiers, the burden on recall could outweigh the advantages gained by direct access. One solution would be to bind auditory descriptions to each identifier that could be played at the user's request without altering the computing environment. However, further study is required to determine the optimal balance of these varied approaches.

RFID-based icons have served as a suitable platform for testing the viability of real-world representations of physical objects. However, the risk of loss and desktop clutter (in the sense that managing a growing stack of objects can quickly become overwhelming) might make this approach less than ideal. Icons fixed across a larger taskbar can alleviate these issues but would eventually limit the amount of file and application bindings. Furthermore, RFID-based icons will always be confined to the number of RFID tags that are available, reducing the freedom of association that is achieved through computer icons and linear access. Utilizing multichannel audio output that has become standard in computing systems might offer a more flexible alternative to the RFID approach. For example, non-speech audio feedback could be combined with text-to-speech output to signal focus change and application selection. While this approach would still reduce auditory overload from text to speech, the benefits of direct access to operating conditions are decreased. The use of computer vision has the potential to accommodate the loss of direct access from an audio-only approach. Hand gestures or printed barcodes could be used to return direct access to the user. When combined with non-speech audio for confirmation, the nonvisual user could reliably select and engage with a larger set of files and applications when compared to RFID-based icons.

8 CONCLUSION

The work presented in this article demonstrates how an audio-based model of computer interaction augmented with tangible representations of semantic information can improve the overall computing experience for blind computer users. Our results indicate three problem areas that tangible interfaces can solve: ephemerality, linear input and output, and unidirectional control.

The ephemerality of sound ensures that audio interfaces that translate visual elements into speech will always be difficult to use. Seeking behavior, memorization, and system reorientation are consistent stumbling blocks that require repetitive, time-consuming actions to resolve. These actions culminate to create a poor user experience fettered with inefficiency. Placing a subset of the visual elements that must be described through speech into a tangible, directly manipulatable object, brings permanence to the audio interface.

Linear interaction models for audio interfaces affect computing performance. Every collection of content that is requested by the user must be accompanied by a semantic description. Headings, links, buttons, and so forth lose their meaning if their intent is not communicated. In an audio interface, this information is presented sequentially with the semantic followed by lexical. While this structure makes it possible to comprehend a complex user interface auditorily, it does so at the cost of efficiency. More words means longer processing times. We address this with our tangible objects by incorporating haptic feedback into the interaction model. For example, the Tangible Scrollbar and the Tangible Taskbar use mechanical resistance to simulate the sensation of ridges or bumps as their slide thumbs are being moved. The resistance is in response to the availability of semantic information as the slide thumb traverses content. This allowed us to remove the semantic information from the audio stream and communicate in at the same time as the lexical information associated with it.

Finally, the traditional graphical user interface is designed to be bidirectional: A user inputs a command through keyboard or mouse, and the state of the system changes visually to match the request. Although this effect exists in audio interfaces, it is not consistent enough to be able to truly think of a screen reader as bidirectional. As our fieldwork demonstrated, mismatches between system state and the user's cognitive model occurred frequently. The only way to recover from a mismatch is to return to the reorientation practices of sequentially processing the environment (i.e., read the list of running applications). The Tangible Desktop starts to address this issue by introducing bidirectional control so when the system state changes, the physical desktop changes as well.

Blind and low-vision communities have been unable to benefit from the advances in user experience that have revolutionized computer interaction for those with sight. While text-to-speech and screen reader software continue to make technology accessible, the limitations imposed by the ephemeral, serialized, and unidirectional constraints of the auditory modality make truly equal access impossible. Grounded through our field work, these limitations point to a design space that has yet to be fully explored. Modern rapid prototyping methods encourage new ways of exploring improvements to the traditional stumbling blocks that blind and low-vision computer users face. Even though many of the challenges we present are often discussed in assistive technology literature, they have yet to be addressed in the manner we prescribe. Our Tangible Desktop system represents one possibility for addressing these limitations, and opens a space for a much broader set of multimodal interactions to be explored.

ACKNOWLEDGMENTS

We thank Ismael Herrera and JoanElaine Anderson at EmpowerTech. This work was supported by the Kliest endowment and is approved by IRB HS2014-1598 at UC Irvine. Finally, the authors thank LouAnne Boyd, Heather Faucett, Sen Hirano, Simone Lanette, and Kate Ringland for thier valuable input on drafts of this paper.

REFERENCES

[1] G. Abner and E. Lahm. 2002. Implementation of assistive technology with students who are visually impaired: Teachers' readiness. *J. Vis. Impair. Blind.* 96, 02 (2002).

[2] Afb.org. 2015. Statistical Snapshots from the American Foundation for the Blind. Retrieved from http://www.afb.org/info/blindness-statistics/2.

[3] Jussi Ängeslevä, Ian Oakley, Stephen Hughes, and Sile OModhrain. 2003. Body mnemonics portable device interaction design concept. *Proceedings of UIST* 3 (2003), 2–5.

[4] C. Asakawa, H. Takagi, S. Ino, and T. Ifukube. 2002. Auditory and tactile interfaces for representing the visual effects on the web. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies.* ACM, 65–72.

[5] O. Bau, Poupyrev, A. Israr, and Harrison. 2010. TeslaTouch: Electrovibration for touch surfaces. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST'10).* 283–292. DOI:http://dx.doi.org/10.1145/1866029.1866074

[6] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and icons: Their structure and common design principles. *Hum.–Comput. Interact.* 4, 1 (1989), 11–44.

[7] 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'10).* ACM, New York, NY., Article 13, 10 pages. DOI:http://dx.doi.org/10.1145/1805986.1806005

[8] Stephen Brewster and Lorna M. Brown. 2004. Tactons: Structured tactile messages for non-visual information display. In *Proceedings of the Australasian User Interface Conference (AUIC'04).* Australian Computer Society, Inc., Darlinghurst, Australia, 15–23.

[9] L. M. Brown, S. A. Brewster, and H. C. Purchase. 2005. A first investigation into the effectiveness of tactons. In *Proceedings of the 1st Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.* IEEE, 167–176. DOI:http://dx.doi.org/10.1109/WHC.2005.6

[10] Laura Dabbish, Gloria Mark, and Víctor M. González. 2011. Why do i keep interrupting myself?: Environment, habit and self-interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 3127–3130. DOI:http://dx.doi.org/10.1145/1978942.1979405

[11] Hilko Donker, Palle Klante, and Peter Gorny. 2002. The design of auditory user interfaces for blind users. In *Proceedings of the 2nd Nordic Conference on Human-Computer Interaction.* 149–155. DOI:http://dx.doi.org/10.1145/572020.572038

[12] W. Keith Edwards, Elizabeth D. Mynatt, and Kathryn Stockton. 1994. Providing access to graphical user interfaces–not graphical screens. In *Proceedings of the First Annual ACM Conference on Assistive Technologies.* ACM, 47–54. http://dl.acm.org/citation.cfm?id=191041, https://smartech.gatech.edu/bitstream/handle/1853/3579/95-35.pdf.

[13] Eelke Folmer and Tony Morelli. 2012. Spatial gestures using a tactile-proprioceptive display. In *Proceedings of the 6th International Conference on Tangible, Embedded and Embodied Interaction (TEI'12).* ACM, New York, NY, 139–142. DOI:http://dx.doi.org/10.1145/2148131.2148161

[14] E. Foulke and T. G. Sticht. 1969. Review of research on the intelligibility and comprehension of accelerated speech. *Psychol. Bull.* 72, 1 (1969), 50.

[15] Victor M. González and Gloria Mark. 2004. "Constant, constant, multi-tasking craziness": Managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 113–120. DOI:http://dx.doi.org/10.1145/985692.985707

[16] Sidhant Gupta, Dan Morris, Shwetak N. Patel, and Desney Tan. 2013. AirWave: Non-contact haptic feedback using air vortex rings. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP'13).* 419–428. DOI:http://dx.doi.org/10.1145/2493432.2493463

[17] James Hollan, Edwin Hutchins, and David Kirsh. 2000. Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.* 7, 2 (2000), 174–196.

[18] Amy Hurst and Jasmine Tobias. 2011. Empowering individuals with do-it-yourself assistive technology. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'11).* ACM, New York, NY, 11–18. DOI:http://dx.doi.org/10.1145/2049536.2049541

[19] Takeo Igarashi, Tomer Moscovich, and John F. Hughes. 2005. As-rigid-as-possible shape manipulation. In *Proceedings of the ACM International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'05).* ACM, New York, NY, 1134–1141. DOI:http://dx.doi.org/10.1145/1186822.1073323

[20] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'97).* ACM, New York, NY, 234–241. DOI:http://dx.doi.org/10.1145/258549.258715

[21] Roberta L. Klatzky and Susan J. Lederman. 2003. Touch. *Handbook of Psychology.*

[22] Susanne Klein, Guy Adams, Fraser Dickin, and Steve Simske. 2013. *3D Printing: When and Where Does It Make Sense?*, Vol. 2013. Society for Imaging Science and Technology, 5–8.

[23] Ravi Kuber, Wai Yu, and M. Sile O'Modhrain. 2010. Tactile web browsing for blind users. *International Workshop on Haptic and Audio Interaction Design.* Springer, 75–84.

[24] Stacey Kuznetsov, Anind K. Dey, and Scott E. Hudson. 2009. The effectiveness of haptic cues as an assistive technology for human memory. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* Vol. 5538. 168–175. DOI:http://dx.doi.org/10.1007/978-3-642-01516-8_12

[25] 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 (May 2007), 247–269. DOI: http://dx.doi.org/10.1080/10447310709336964

[26] Jennifer Mankoff, Holly Fait, and Tu Tran. 2005. Is your web page accessible?: A comparative study of methods for assessing web page accessibility for the blind. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'05).* ACM, New York, NY, 41–50. DOI: http://dx.doi.org/10.1145/1054972.1054979

[27] Gloria Mark, Shamsi T. Iqbal, Mary Czerwinski, and Paul Johns. 2014. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 3025–3034. DOI: http://dx.doi.org/10.1145/2556288.2557204

[28] Elizabeth Mynatt. 1997. Transforming graphical interfaces into auditory interfaces for blind users. *Hum.-Comput. Interact.* 12, 1 (1997), 7–45. DOI: http://dx.doi.org/10.1207/s15327051hci1201&2_2

[29] Elizabeth D. Mynatt and Gerhard Weber. 1994. Nonvisual presentation of graphical user interfaces: Contrasting two approaches. In *Proceedings of the SIGCHI Conference on Computer Human Interaction (CHI'94).* ACM, New York, NY, 166–172. DOI: http://dx.doi.org/10.1145/191666.191732

[30] J Nielsen. 1994. Heuristic evaluation. *Usabil. Inspect. Methods* 17, 1 (1994), 25–62.

[31] Peter Plessers, S. Casteleyn, Yesilada Yesilada, O. D. Troyer, R. Stevens, S. Harper, and C. Goble. 2005. Accessibility: A web engineering approach. In *Proceedings of the 14th International Conference on World Wide Web.* ACM, 353–362. DOI: http://dx.doi.org/10.1145/1060745.1060799

[32] L. H. D. Poll and R. P. Waterham. 1995. Graphical user interfaces and visually disabled users. *IEEE Trans. Rehabil. Eng.* 3, 1 (Mar.1995), 65–69. DOI: http://dx.doi.org/10.1109/86.372894

[33] Denise Prescher, Gerhard Weber, and Martin Spindler. 2010. A tactile windowing system for blind users. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'10).* 91–98. DOI: http://dx.doi.org/10.1145/1878803.1878821

[34] T. V. Raman. 1996. Emacspeak—A speech interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'96).* ACM, New York, NY, 66–71. DOI: http://dx.doi.org/10.1145/238386.238405

[35] Dan Ratanasit and Melody M. Moore. 2005. Representing graphical user interfaces with sound: A review of approaches. *J. Vis. Impair. Blind.* 99, 2 (2005), 69.

[36] John T. Richards, Kyle Montague, and Vicki L. Hanson. 2012. Web accessibility as a side effect. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'12).* ACM, New York, NY, 79–86. DOI: http://dx.doi.org/10.1145/2384916.2384931

[37] Brigitte Röder, Frank Rösler, and Helen J. Neville. 2001. Auditory memory in congenitally blind adults: A behavioral-electrophysiological investigation. *Cogn. Brain Res.* 11, 2 (2001), 289–303. DOI: http://dx.doi.org/10.1016/S0926-6410(01)00002-7

[38] W. Sapp and P. Hatlen. 2010. The expanded core curriculum: Where we have been, where we are going, and how we can get there. *J. Vis. Impair. Blind.* 104, 6 (2010), 338.

[39] Valkyrie Savage, Xiaohan Zhang, and Björn Hartmann. 2012. Midas: Fabricating custom capacitive touch sensors to prototype interactive objects. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST'12).* ACM, New York, NY, 579–588. DOI: http://dx.doi.org/10.1145/2380116.2380189

[40] Anthony Savidis, Constantine Stephanidis, Andreas Korte, Kai Crispien, and Klaus Fellbaum. 1996. A generic direct-manipulation 3d-auditory environment for hierarchical navigation in non-visual interaction. In *Proceedings of the 2nd Annual ACM Conference on Assistive Technologies (ASSETS'96).* ACM, New York, NY, 117–123. DOI: http://dx.doi.org/10.1145/228347.228366

[41] Chris Schmandt. 1998. Audio hallway: A virtual acoustic environment for browsing. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology.* ACM, 163–170.

[42] Bertrand Schneider, James M. Wallace, Paulo Blikstein, and Roy Pea. 2013. Preparing for future learning with a tangible user interface: The case of neuroscience. *IEEE Trans. Learn. Technol.* 6, 2 (2013), 117–129.

[43] Section508.gov. 1998. Section 508 Law and Related Laws and Policies. Retrieved from https://www.section508.gov/content/learn/laws-and-policies.

[44] Orit Shaer and Eva Hornecker. 2010. Tangible user interfaces: Past, present, and future directions. *Found. Trends Hum.-Comput. Interact.* 3, 12 (2010), 1–137. DOI: http://dx.doi.org/10.1561/1100000026

[45] R. Sodhi, Poupyrev, M. Glisson, and A. Israr. 2013. AIREAL: Interactive tactile experiences in free air. *ACM Trans. Graph.* 32, 4 (2013), 134. DOI: http://dx.doi.org/10.1145/2461912.2462007

[46] Markel Vigo and Simon Harper. 2013. Challenging information foraging theory: Screen reader users are not always driven by information scent. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media (HT'13).* ACM, New York, NY, 60–68. DOI: http://dx.doi.org/10.1145/2481492.2481499

[47] Markel Vigo and Simon Harper. 2014. A snapshot of the first encounters of visually disabled users with the Web. *Comput. Hum. Behav.* 34 (2014), 203–212. DOI: http://dx.doi.org/10.1016/j.chb.2014.01.045

[48] Holly S. Vitense, Julie A. Jacko, and V. Kathlene Emery. 2002. Multimodal feedback: Establishing a performance baseline for improved access by individuals with visual impairments. In *Proceedings of the 5th International ACM Conference on Assistive Technologies.* ACM, 49–56.

[49] W3C. 2014. Web Accessibility Initiative (WAI). Retrieved from http://www.w3.org/WAI/.

[50] W3C. 2014. Web Content Accessibility Guidelines (WCAG). Retrieved from https://www.w3.org/WAI/intro/wcag.

[51] Webaim.org. 2016. Web Accessibility in Mind. Retrieved from http://webaim.org/.

[52] Malte Weiss, Simon Voelker, Jan Borchers, and Chat Wacharamanotham. 2011. FingerFlux: Near-surface haptic feedback on tabletops. In *Proceedings of the 24th Symposium on User Interface Software and Technology (UIST'11).* 615–620. DOI: http://dx.doi.org/10.1145/2047196.2047277

[53] Malte Weiss, Julie Wagner, Yvonne Jansen, Roger Jennings, Ramsin Khoshabeh, James D. Hollan, and Jan Borchers. 2009. SLAP widgets: Bridging the gap between virtual and physical controls on tabletops. In *Proceedings of the SIGCHI Conference on Computer Human Interaction (CHI'09).* ACM, New York, NY, 481–490. DOI: http://dx.doi.org/10.1145/1518701.1518779

[54] Brian Wentz, Harry Hochheiser, and Jonathan Lazar. 2013. A survey of blind users on the usability of email applications. *Univers. Access Inf. Soc.* 12, 3 (Aug. 2013), 327–336. DOI: http://dx.doi.org/10.1007/s10209-012-0285-9

[55] Cheng Xu, Ali Israr, Ivan Poupyrev, Olivier Bau, and Chris Harrison. 2011. Tactile display for the visually impaired using TeslaTouch. In *Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems (CHIEA'11).* ACM Press, New York, New York, 317. DOI: http://dx.doi.org/10.1145/1979742.1979705

[56] Yeliz Yesilada, Simon Harper, Carole Goble, and Robert Stevens. 2004. *Screen Readers Cannot See.* Springer, 445–458.

[57] Jiajie Zhang and Vimla L. Patel. 2006. Distributed cognition, representation, and affordance. *Pragmat. Cogn.* 14, 2 (2006), 333–341.

[58] Li Zhou, Amy T. Parker, Derrick W. Smith, and Nora Griffin-Shirley. 2011. Assistive technology for students with visual impairments: Challenges and needs in teachers' preparation programs and practice. *J. Vis. Impair. Blind.* 105, 4 (2011), 197.

[59] Oren Zuckerman and Ayelet Gal-Oz. 2013. To TUI or not to TUI: Evaluating performance and preference in tangible vs. graphical user interfaces. *Int. J. Hum. Comput. Stud.* 71, 7–8 (2013), 803–820. DOI: http://dx.doi.org/10.1016/j.ijhcs.2013.04.003

Received April 2016; revised March 2017; accepted March 2017