now

the essence of knowledge

Crowdsourcing Accessibility: Human-Powered Access Technologies

Erin Brady
Department of Human-Centered Computing
School of Informatics and Computing
Indiana University – Purdue University Indianapolis
brady@iupui.edu

Jeffrey P. Bigham
Human-Computer Interaction Institute
Carnegie Mellon University
jbigham@cmu.edu

Contents

Abstract

People with disabilities have always engaged the people around them in order to circumvent inaccessible situations, allowing them to live more independently and get things done in their everyday lives. Increasing connectivity is allowing this approach to be extended to wherever and whenever it is needed. Technology can leverage this human work force to accomplish tasks beyond the capabilities of computers, increasing how accessible the world is for people with disabilities. This article outlines the growth of online human support, outlines a number of projects in this space, and presents a set of challenges and opportunities for this work going forward.

1

Introduction

People with disabilities have always leveraged the assistance of people around them to help them get things done. Supporters are drawn from family members, friends, employees, or nearby strangers. This can work well when someone is available, but it breaks down if a supporter isn't there when needed. Constant access to support has always been possible for those who could afford it, but the cost may quickly grow prohibitive because the supporter needs to be available at all times even if they are only occasionally needed.

Technology has changed this tradeoff because in the past few decades the people who provide assistance no longer need to be physically proximate for many critical tasks or services. This has opened a rich new era in technology design —creating tools that can incorporate human intelligence when and how it is needed to flexibly support people with disabilities, rather than relying on machine intelligence alone, which may be incomplete or limited to specific situations.

These crowd-powered systems, as they have come to be called [Bernstein, 2012, Lasecki et al., 2014b], address difficult problems, many of which are of particular and practical interest to people with disabilities, by combining human intelligence and computation in new ways. People

Figure 1.1: A diagram of the Faxcess system from the early 1990s for helping blind people interpret visual information. In this example, a blind person has taken a picture of the box for a frozen dinner with a digital scanner, and sent it to a human supporter who is reading the information to him over the phone. It is interesting to note that the supporter herself has a physical disability. While this was not called human computation or crowdsourcing at the time, it presaged many of these ideas.

with disabilities are some of the first true users of this technology, so understanding how they use it and manage their expectations around it can be more broadly informative. For instance, nearly 10,000 blind users have asked VizWiz a question about an image that they took, giving us insights not only into what visual information blind people would like to know about but also how they approach asking workers online for answers about what is around them.

In order to understand how to design these types of technologies, we draw from several successful systems utilizing the paradigm of *human-powered access technologies*. In this article, we first trace how people with disabilities went from being early adopters of technology for re-

mote support from other people to becoming some of the first users of technology powered by the crowd. We discuss methods used not only to obtain high-quality, reliable work with these systems, but also how people with disabilities have actively managed the work and workers to ensure that their expectations were met. We discuss who the people in the crowd are, how they are recruited, what incentives they receive, the ethics of crowdsourcing, and what connections (if any) they have to the people that they support. We explore these ideas through a set of examples of technology in this space, and then describe three case studies of human-powered access technologies, which we analyze through the lens of our design dimensions. The first, Social Accessibility, allows people with disabilities to request accessibility meta-information about webpages, which can then be reused by others visiting the website after them. The second case study describes a technology that we have developed called VizWiz that answers visual questions for blind people in nearly-realtime by sending them to sighted crowd workers. The third case study covers Scribe, a real-time captioning tool that combines novice transcriptions from multiple workers into a coherent stream of captioning. From these prior examples, we extract design dimensions that we believe can help characterize technology in this space, provide insights to designers of new technology so that their work can build on what has come before, and start a framework to allow new technologies to be compared and contrasted to prior work. We then conclude with a list of remaining challenges for work in this area.

1.1 Human-Powered Access Technology

The term human-powered access technology is intended to capture the broad range of technologies supporting access for people with disabilities by leveraging human intelligence, effort, and perception [Bigham et al., 2011]. This term is intended to cover a variety of different kinds of systems powered by humans that support people with disabilities by making something accessible that was not previously. Many of these systems include substantial computational aspects, but we purposefully exclude those technologies that rely only on computation.

Access technologies have often included human intervention to make content accessible to people. Initially, this was due to a lack of existing technologies to solve access problems. For instance, human readers were used by blind people to access text before optical character recognition (OCR) tools and audio books became common. While automated tools now exist to aid users with these access problems, many are in early stages and do not function perfectly. OCR works well for clear text, but does not perform well on handwriting; automatic speech recognition (ASR) similarly works well in ideal conditions if it's been trained for a particular speaker, but is not sufficient for recognizing noisy fragments of speech. Because of these limitations, humans are often used to train, supplement, or replace automated solutions, as many tasks that are hard for a computer to do, *e.g.,* visual tasks, recognizing and understanding human speech, are comparatively easy for people.

Human computation and crowdsourcing are two related terms used to discuss work that is performed by people and mediated by technology [Quinn and Bederson, 2011]. While the two concepts have some overlap, they take advantage of human labor in distinct ways. In instances of human computation, people are used to perform tasks too difficult for a computer to do, often in conjuction with the computer. One well-known example of this is CAPTCHAs, the "Completely Automated Public Turing Test to Tell Computers and Humans Apart" [Von Ahn et al., 2004], which use AI-hard problems that are easy for humans to solve as a way to distinguish between real users and automated bots. Two facets typically define human computation tasks: tasks are structured in a way that, if technology "caught up" and could solve them easily, the human labor could be replaced by computers; and a computational structure exists that distributes the tasks and aggregates the results intelligently [Quinn and Bederson, 2011].

Crowdsourcing is a paradigm of distributing jobs, which traditionally may have been completed by a fixed employee, to remote people online. While similar in nature to human computation, crowdsourcing is not limited to tasks that are hard for computers to do automatically, and in fact can serve as a replacement for technological solutions if they are expensive or unwieldy [Quinn and Bederson, 2011].

Though some tasks may have a single component and need only one worker, many workflows have been developed to allow crowdsourcing to be performed with multiple workers simultaneously or in sequence, allowing them to supplement or exceed a single worker's output [Little et al., 2010]. Crowdsourcing is also related to collective intelligence. Generally, for human-backed access technologies we are interested in "directed crowdsourcing", in which someone, *e.g.,* the person with a disability, commissions the work to be done.

Human computation lets users access information that may not yet be accessible by technology, while crowdsourcing enables fast access to humans who can provide assistance. This article describes the advent of technologies that leverage the power of humans recruited to improve the support available for people with disabilities. Our focus is on crowdsourcing systems, those that rely on open calls to pools of workers, because this seems to be one of the better ways of recruiting workers improving accessibility support, but we include discussion of systems across the range of human-powered access technology to help fill out the space.

In the next sections, we discuss more about *people with disabilities* —the history of disability, accessibility, and access technologies —and *the crowd* —the people within it, their motivations, and how crowds composed of different people can be used for different purposes.

2

Disability, Accessibility and Access Technologies

Understanding access technologies is fundamentally dependent on understanding disability. The way disability is defined and addressed has varied dramatically over time and across cultures, due in large part to the emergence of the disability rights movement, and has resulted in changing perspectives of accessibility and the design of access technologies. In this chapter, we present multiple definitions and models of disability, and discuss the history of accessibility initiatives and assistive tools that have influenced the adoption of human-powered access technologies.

2.1 Definitions and Models of Disability

The World Health Organization's International Classification of Functioning, Disability and Health (ICF) defines disability as any occurrence or combination of *impairments, activity limitations or participation restrictions*. The ICF provides definitions for these terms [ICF, 2002]:

- 'Impairments are problems in body function or structure such as a significant deviation or loss.'

- 'Activity Limitations are difficulties an individual may have in executing activities [the execution of a task or action].'

- 'Participation Restrictions are problems an individual may experience in involvement in life situations.'

Disability can be thought of as a combination of physical factors of the individual (bodily impairments) and social or environmental factors (*e.g.,* design of physical objects, social attitudes) which work together to limit activity and participation in daily life. These factors are institutionalized in two prominent models of disability: the *medical model*, which focuses on impairment as the cause of disability, and the *social model*, which focuses on attitudes and designs which disable people who do not have certain abilities.

The medical model of disability views the bodily impairment of a person with a disability as a problem which should be fixed through medical intervention [ICF, 2002]. This model is very focused on the individual and their physical impairment. Rather than examining how people are disabled by a lack of access to resources, it emphasizes rehabilitation and receiving care from others [Barnes and Mercer, 2006]. This focus on the individual's impairment often leads to language of 'deficit' being used in medical writing, discussing abilities which a person with a disability lacks when compared to a person without that disability.

The social model of disability instead focuses on the individual's physical and social environment as being disabling. In the social model, a person's disability stems from cultural denials of resources or exclusion [ICF, 2002], a view which separates disability from any specific bodily impairment [Shakespeare and Watson, 2001]. Social exclusion and inaccessibility of services are the reason that people are disabled in certain situations, and changes in access can change a person's disability status [Mankoff et al., 2010]. This model of disability dictates changes in political and public perceptions of people with disabilities without requiring the individual to change themselves, overlapping with

goals of the disability rights movement [Charlton, 1998]. The social model has had political implications, provided instrumental steps to improve accessibility, and has psychological benefits for people with disabilities [Shakespeare and Watson, 2001].

The ICF combines these two models into a third, the *biopsychosocial model*, which holistically considers physical impairments and the individual perspective with the impacts of the community and society that people with disabilities live amongst [ICF, 2002]. This combined approach addresses the medical model's single-track focus on impairment and individual deficit (by incorporating sociological analyses) and the social model's minimization of physical impairments (by acknowledging physical pain or other negative side effects of impairments) [Shakespeare and Watson, 2001].

These models are inherently tied into the language used to describe people with disabilities. Medical models often use the language of impairment or deficit (*e.g.,* a hearing-impaired person), while social models often use person-first language to emphasize the person before their impairment (*e.g.,* a person with vision loss). These language choices are often debated, and preferred styles change over time and across cultures, so careful attention to language is important in reading and writing about people with disabilities to avoid stigmatizion or oppression groups [Cavender et al., 2008, Mankoff et al., 2010].

2.2   A History of Accessibility

Within the United States[1], services for and attitudes towards people with disabilities have changed radically within the last century. Many of these cultural and legislative changes were triggered by war efforts - the emergence of prosthetics and manual wheelchairs began after the Civil War, but rehabilitation efforts for veterans were not codified until the Soldiers Rehabilitation Act of 1918, followed shortly by the Vocational Rehabilitation Act of 1920 for general audiences [Scherer, 1996].

---

[1]In this section, we focus on the history of accessibility within the United States, where both authors are located; however, we recognize that conceptions of disability change dramatically across cultures, and that many countries have had similar, but unique, trajectories of rights and inclusion for people with disabilities.

These legislative initiatives addressed problems of disability, but in a collective sense. The individual's needs and abilities were not considered; instead any person using a wheelchair was considered a 'wheelchair user', and was given a general-purpose device and rehabilitation plan. The change towards individualized services, a move characterized by Scherer [1996] as shifting the focus from all 'people' to each 'person', occurred slowly, and gained traction through the disability rights movement of the late 1960s and 1970s. This movement, which focused on independent living and personalized services, lead to the Rehabilitation Act of 1973 guaranteeing protections in the federal context for people with disabilities, and the Americans with Disabilities Act (ADA) of 1990, which more broadly prohibits discrimination against people with disabilities.

Many scholars have noted that this movement towards equal rights for people with disabilities occurred much later than most other civil rights movements in the US, after successes of the women's liberation movement of the 1910s and the African-American civil rights movement of the 1950s [Charlton, 1998]. As such, much of the effective legislation has been drafted within the last fifty years, and it often addresses technological advances explicitly. Section 508[2], a 1998 amendment to the Rehabilitation Act of 1973, ensures access to federal electronic and information technologies by both employees and members of the public with disabilities.

2.3   Assistive and Access Technologies

A wide variety of assistive and access technologies have been developed for people with disabilities. Initial tools were manual, while later tools incorporated automatic technologies or human involvement. We discuss human involvement in assistive devices further in Chapter 4, but here chart the general history of manual and automatic tools, and discuss design methods being utilized by people building new assistive technologies.

---

[2]29 U.S. Code §794d - Electronic and information technology

These technologies can be used for a variety of purposes in daily life. Galvin and Scherer [1996] proposed a categorization of assistive device purposes (personal/self-care, mobility, transportation aids, communication devices, and specific technologies for people who have visual impairments, are hard of hearing or Deaf) and contexts (leisure, play, home and work).

The differences between *assistive* technologies and *access* technologies are nuanced; an assistive technology may providing unsolicited help or interpretation of something to a user with a disability, while an access technology provides a direct translation from one sensory output to another, taking something that is inaccessible and making it accessible so that the user can choose how to interact with it. For example, a wheelchair functions as an access technology, providing access mobility to a person with a mobility impairment, while images on the internet which have been given alternative text by the developer are assistive technologies.

Early manual access technologies were mostly rehabilitative, including manual wheelchairs (first introduced to the general public in the late 1800s) and simple prosthetics. Other manual technologies, like 'reacher' devices which extend the arm range of someone who may have limited mobility, have followed, and are still widely used due to their practicality, low cost, and ease of replacement [Scherer, 1996].

Technological advances allowed these early devices to be improved upon. Wheelchairs are now often battery-powered, reducing the physical exertion required for mobility. People with communication disabilities historically used manual 'letter boards' to communicate, pointing at each letter of a word in order and then waiting for observers to piece the letters together into full words; now, text-to-speech (TTS) engines allow people to type words which are spoken aloud by a computer, and computerized augmented and alternative communication (AAC) tools let users select images or phrases to form messages [Scherer, 1996]. The addition of technology to manual tools has made them more flexible and convenient for users.

Many of the early assistive technologies were human-powered. Telecommunications relay services used human operators to dictate

text messages from a Deaf caller to a hearing caller, and people with visual impairments often hired readers to speak aloud text to them. While some assistive technologies are manual tools enhanced by existing technologies or human assistants, other technological advances have been spurred by a drive to create more accessible technologies. Technologies like text-to-speech and optical character recognition were driven and perfected for the Kurzweil Reader, an early text access tool for people with visual impairments [Kurzweil, 2003], and speech-to-text interfaces for computers were for some time only used by people with disabilities, seeing little adoption in the general population [Shneiderman, 2000]. Because of these strong ties to early tools, assistive technologies have been an area of focus within HCI research since its conception.

Methodologies have also been developed to help HCI researchers and designers create effective assistive technologies, mostly by including users with disabilities into the design process and considering the various ranges of abilities that a set of users may possess. Universal design [Story, 1998] and universal usability [Vanderheiden, 2000] are founded around making technologies accessible for any user, regardless of any disability they may or may not have. These design approaches may provide optimal results in terms of being usable for as many user groups as possible, but in practice many researchers and designers found universal usability difficult to implement.

As a result, user-sensitive inclusive design [Newell et al., 2011] was proposed to relax the 'universal' aspect and promote more involvement of people with disabilities in the early stages of design, and ability-based design [Wobbrock et al., 2011] proposed that interfaces be designed to work in different ways to suit the specific abilities of any user, rather than focusing on what a user with a disability would be unable to do. These theories and principles for design have pushed assistive technology designers and researchers to broaden their conceptions of disability and ability, and include users with disabilities more frequently in their processes.

3

Crowdsourcing and Remote Workers

While crowdsourcing can be viewed as a "black box", where tasks are posted, completed and returned some indefinite amount of time later to the requester, a more nuanced understanding of crowdsourcing is generally necessary to make effective use of it. The crowds in crowd-powered systems can be composed of a vast number and variety of people, and recognizing their abilities, limitations, and motivations is crucial for choosing the work to be performed and designing the interface that workers use. Bigham et al. [2014] offer a general overview of different types of crowd work and workers.

In this section, we discuss a number of factors which may impact how crowdsourcing is used for access technologies: the composition of the crowd of workers, the kind of work they are able to do, and the incentives they receive. By understanding the crowd less as an abstract source of work, and more as a group of individuals, we can gain a greater understanding of why these systems work. In this section, we discuss crowdsourcing for any type of task; later in the article our focus narrows to accessibility-related tasks.

3.1   Composition of the Crowd

In this section, we discuss several different kinds of crowds, which may be particular useful for identifying the right people to power a specific human-powered access technology.

3.1.1   Crowdsourcing Marketplaces

Many systems are powered by paid workers on crowdsourcing marketplaces like Amazon's Mechanical Turk. As Mechanical Turk has been used extensively by the research community, the demographics are among the most well-studied in crowdsourcing, and can provide an initial grounding of the composition of crowds on pay-for-work sites.

The Mechanical Turk Platform

Requesters, the task creators who are looking for workers from the platform, are able to post their human intelligence tasks (HITs) on the site in many ways:

- as individual, one-off tasks (*e.g.,* a request for information),

- as a large group of similar tasks where one worker could answer all of them (*e.g.,* a large set of images that need to be described), or

- as a large group of identical tasks that all should be completed by different workers (*e.g.,* a survey that needs multiple participants).

Requesters are able to choose from a set of pre-determined, frequent task types – including surveys, image categorization, or transcription – or design their own task. They can then select how much to pay workers for each completed task, how long the worker has to finish the task once they have started, and who will qualify to work on the task. The requester can use the default Masters qualification (where users must have completed a large number of HITs with a minimum approval rate), or set their own qualification types based on HITs completed, average approval rate, or worker location.

Figure 3.1: The Mechanical Turk interface for turkers

The Mechanical Turk workers, frequently known as "Turkers", can then search through and choose HITs to work on from the web interface (Figure 3.1). This interface shows HITs in reverse chronological order of their posting, so the most recently posted tasks are always on the front page. Turkers can search for HITs based on the amount they pay, the qualifications required, or by using keywords for the types of tasks that they prefer to work on.

Once a turker has found a HIT they may be interested in, they can click to view the task before accepting it. They are able to see a live preview of the HIT, and can decide whether they want to work on it or not (though for some HITs with multiple pages, this may not fully represent the amount of work they will have to do). Once a task is accepted, the worker has a certain amount of time to complete the HIT, or return it if they cannot complete it.

Once work is returned, the requester can approve or reject the work based on quality. After a specified period, work will be auto-approved if the requester has not been able to review it. The requester can provide feedback if a task was rejected, or grant a bonus payment if work was exceptional. Requesters can also block turkers if they are consistently performing poorly, which means the worker will no longer be able to accept their HITs.

While Mechanical Turk itself does not enable turkers to rank requesters, other sites like Turkopticon [Irani and Silberman, 2013] and Turker Nation [Martin et al., 2014] have arisen to fill that need. Turkers can rank requesters on many characteristics, such as responsiveness and how well they pay or review work, and then are able to use those rankings to be more informed while choosing HITs.

Turkers

Turkers, the workers on Mechanical Turk, come from many different backgrounds, and participate on the platform for a variety of reasons. Mechanical Turk has been active since 2005, and the demographics of turkers has naturally changed and fluctuated since then.

When Mechanical Turk first started, only turkers from the United States (with a US-based bank account) were able to receive cash payments for their work, while other turkers had to accept an Amazon gift card instead [Ipeirotis, 2008]. As a result, the majority of turkers (more than 75%) were from the US during the early years of the service. These turkers were often highly educated (with a bachelors degree or higher), and had other sources of income which their Mechanical Turk income was supplementing.

In 2007, Indian Rupees were added as another form of payment that could be received directly, rather than via gift card [Ipeirotis, 2010]. This had a significant impact on the geographic diversity of turkers participating on the platform - by 2010, US-based turkers composed 47% of available turkers, while Indian-based turkers made up 35%. Ross et al. [2010] performed an analysis of the changes in Turker demographics over time, finding that the percentage of Indian turkers on the platform had increases drastically, while an initial gender imbal-

ance favoring women had evened out so that the platform was nearly gender-equal, and the average age of turkers had begun to drop.

The highly-international nature of Mechanical Turk presents both opportunities and challenges for requesters. Many turkers on the platform from other countries are multilingual in at least their own native language and English, meaning that language translation tasks can be completed for certain languages easily [Pavlick et al., 2014]. However, problems may arise from the site's complicated interface, or the differences in cultural context between the (primarily US-based) requesters and Indian turkers [Khanna et al., 2010].

Turkers may participate in crowd work for a variety of reasons. Early perceptions of the platform's userbase consisting mostly of stay-at-home moms or others trying to make a supplemental income [Ross et al., 2010] has given way to a more nuanced view of the turkers, who may use the platform as a financial safety net [Martin et al., 2014] or, in the case of turkers with disabilities, prefer its flexibility and affordances over a typical office job [Zyskowski et al., 2015].

3.1.2 Other Crowdsourcing and Remote Work Platforms

Mechanical Turk has been a popular crowdsourcing platform for academic work because it was the first platform of its type and is reasonably robust to different kinds of jobs. Many similar platforms are also in the space with a variety of differences. Many of these platforms have specific methods of crowdsourcing, target certain groups as workers, or allow only certain types of work.

Many crowdsourcing platforms recruit workers who are disadvantaged as a way to accomplish work while benefiting social good. Lead-Genius[1] (formerly MobileWorks) is a sales technology platform with a workerbase composed of women, youth, and people who are unemployed, while CrowdFactory is a similar business-oriented group that works with people in the developing world, and Samasource is a non-profit organization which works with many of the same groups [2]. These platforms aim to connect people who cannot perform typical jobs (due

---

[1]`http://www.leadgenius.com`
[2]`http://www.samasource.org`

to family responsibilities or social forces) with tasks that they can complete without disrupting their current lifestyles.

The previously mentioned sites all gear themselves towards microwork - small, standalone tasks that can be completed by anyone. Other platforms are built to facilitate longer-term relationships instead. oDesk and Elance allow individuals to hire freelance workers for specific tasks which may take time and repeated iterations to complete [Horton, 2010]. Requesters post jobs on the site, which workers can bid on. The workers host portfolios on the site, so requesters can determine if the worker's previous work exhbits an appropriate level of experience.

3.1.3 Private Crowds

Many companies host their own private crowds that can be used for their own purposes, *e.g.,* URHS at Microsoft [Wang et al., 2012]. Companies can choose from a plethora of different "crowd" marketplaces, from Amazon Mechanical Turk to those much more similar to traditional outsourcing firms [Horton, 2010]. For instance, a large number of firms have started to specialize in crowdsourced transcription [Zhou et al., 2013]. Even Mechanical Turk started as an internal crowd that Amazon used to help improve its product listings. Less has been written about these private corporate crowds as they are generally internal to the corporations that host them. For a variety of reasons, companies are often hesitant to talk about them.

Private crowds have had an active history in human-powered access technology for at least two reasons. The first is that many of the tasks that people are asked to do involve sensitive material that people with disabilities would want to keep private. One of the earliest examples of what we might call crowdsourcing today are the relay services used by Deaf and hard of hearing people to make phone calls. The deaf person would type (via a TTY device or later via IM), and a person at the relay office would convey the information to a hearing person. Today, these often operate using video links where the Deaf user whose preferred language is ASL signs to the relay interpreter. Relay services are generally governed by strict confidentiality rules.

The second reason that many older services used private crowds is that alternative general purpose crowdsourcing platforms were available after the uptake of private crowds. Crowdsourcing-like activities have existed at organizations long before the term "crowdsourcing" was coined, and so another difficulty in tracing this history is that companies did not (and still do not) necessarily consider what they do to be crowdsourcing.

3.1.4   Other Forms of Online Group Work

Beyond crowdsourcing, other forms of group work also occur online, often under motivations of altruism or entertainment. Two of the most well-studied examples are citizen science and games with a purpose.

Citizen science allows individuals without scientific expertise to contribute to scientific goals. Individual people can serve as environmental sensors to collect data from remote locations (with each person recording local weather conditions or occurrences of wildlife), or as a first-round of human screeners for large data processing tasks that are too high-volume for scientists to complete on their own. Sites like Zooniverse[3] allow participants to examine the sky for undiscovered galaxies, or to annotate historical documents.

Games with a Purpose also allow people online to complete tasks, with the added benefit of being fun and engaging for the player. These games often have a social component, allowing players to compete against each other while simultaneously completing human computation. The ESP Game, described in the next chapter, is an example of a game that benefits people with disabilites by generating descriptions of the contents of an image.

3.1.5   Ethics and Crowdsourcing

As the use of crowdsourcing has grown, concerns have been raised about the ethics of crowd work. Some of these concerns have to do with the employment conditions of crowd workers on paid marketplaces, who are often paid little and lack fair ways to handle dis-

---

[3]http://www.zooniverse.org

putes. This has led to workers banding together to explore what sort of collective action might make sense in the context of crowdsourcing, *e.g.,* creating a worker bill of rights ([Salehi et al., 2015], available at `http://www.wearedynamo.org`). Other concerns revolve around what kind of *future* crowd work may lead to for a worker, and how we might design systems so that crowd workers can benefit over time by gaining expertise and credentials ([Kittur et al., 2013]). A favorable aspect of crowd work in the context of human-powered access technology is that most people will agree that the end result is a positive (the technology is designed to do *good*). Even in this context, we might be concerned about what happens when there are too many applications that require too much of everyone's time to power, or when someone contributes too much of their own time to helping out of a desire to do good without adequate compensation.

3.2   Designing Crowd Work

On many crowdsourcing platforms, tasks are posted with the assumption that workers are novices, with little domain expertise. While this is not always the case (especially for freelancing sites like oDesk), the assumption has many implications for how tasks are designed, how workers should be trained, and how work quality can be verified.

3.2.1   Task Design

Tasks can be designed specifically for domain experts in a certain area, or more generally for users of any skill level to complete. When tasks are intended for domain experts, they often are first asked to complete a qualification test, which allows them to demonstrate their skills and prove that they know enough to complete the tasks correctly. These domain-specific tasks can require a variety of expertise, from fluency in a certain language to perform translations [Pavlick et al., 2014] to photograph or video editing skills on freelancing sites.

For tasks which allow workers of any level to participate, they must be structured so that a newcomer can understand how to perform tasks quickly. Many tasks included examples of how to perform the work

(or training, as discussed below). Other requesters split large, complex tasks into smaller steps which may be more amenable to novice workers.

A prominent example of this is Soylent [Bernstein et al., 2010], which introduced a *Find-Fix-Verify* model for document proofreading tasks. Rather than asking a single worker to proofread and edit a full paragraph, the editing is split into three tasks: finding problems in the paragraph, correcting those problems, and verifying the correction. By splitting tasks down in this way, novice workers are given less to learn how to do, and their work undergoes review by other turkers in the subsequent step to ensure that bad corrections are excluded.

Other systems, like TurKit [Little et al., 2010], also utilize incremental tasks to help workers succeed while optimizing the financial costs of the system. TurKit allows for *crash-and-rerun* programming, in which tasks are frequently stopped to wait for incoming data, and then restarted once sufficient data is collected. This model can be used for iterative writing tasks (where subsequent users iteratively improve writing samples, and vote on the quality of other turker's improvements), or for blurry or messy handwriting recognition (where an initial user provides their best guess, and subsequent workers make improvements as they can).

A different type of task design can be used for projects which require specialized knowledge but little training. In the GalazyZoo project, where citizen scientists classify images of galaxies, the scientists may come in knowing little about astronomy. Rather than asking participants to take part in training to learn about different types of galaxies, the task is split into small, easy to understand portions which are triggered hierarchically - for example, if the answer to the first question about the galaxy's shape is "round", the second question asks for the round shape (oval, circular, or elongated), while if the answer for the first question is "features or disk", the following questions ask for details about the features, or characteristics of the disk's arms and center. This task design allows the scientists to collect detailed information without forcing the workers to memorize and be able to recall all the different galaxy types.

3.2.2   Worker Training

While incremental or hierarchical task design can overcome a lack of domain expertise for some tasks, others require some initial training for the worker to understand how to complete them. However, with speed as a crucial factor to many crowdsourcing algorithms, training must be done in advance through qualification or training HITs, or occur quickly during the completion of the tasks.

Many Mechanical Turk tasks provide instructional steps at the beginning, or make workers to watch a training video before the task begins. Other projects have more interactive instructions. Operation War Diary, a citizen science project run by the same group as Galaxy-Zoo, requires volunteers to perform a 10-minute tutorial, in which they actually perform all the types of data annotation tasks they will complete for the real tasks on a sample, and are given immediate feedback on whether they performed correctly or not. Both researchers and requesters have begun to focus on ways to "shepherd" the crowd, providing real-time feedback which allows workers to improve their performance and complete tasks better than when they are not guided [Dow et al., 2012].

3.2.3   Quality Verification

While the work being performed on crowdsourcing platforms can be collected in massively large scale, review of all of the data collected by workers can take nearly as long as production, and work review can become a significant bottle neck for requesters to complete their tasks [Ipeirotis et al., 2010]. While most turkers have a strong sense of ethics when it comes to their work [Martin et al., 2014], a few workers looking to make money with little effort can have significant ramifications for requesters.

One standard method of quality verification is "gold-standard" questions, where the task includes a question or piece of work with a known answer [Downs et al., 2010]. Gold standards can range from questions with answers that most humans would know to prevent bots (*e.g.,* "$2 + 2 = ?$"), or questions which provide the answer inline to test

worker attention (*e.g.,* "The correct answer to this question is choice B"). These questions can serve as an initial screening to eliminate low-quality work, leaving only work where these questions were answered correctly to go through further verification.

Other methods of work verification include consensus methods (where the same task is completed by multiple workers, and only work that agrees with the majority vote is accepted) or using machine learning to try to determine the quality of the individual worker or their work [Ipeirotis et al., 2010]. As discussed above, many systems rely on turkers to verify the quality of previously completed work. However, this adds to the expenses of the project, and is vulnerable to the same problems - workers or bots may just auto-approve others' work.

3.3 Incentives for Crowd Work

Workers may have many different motivations for participating in crowd work. These motivations may work in concert with each other, and have impacts on the types of tasks workers choose to do, the quality of work they provide, and how long they remain engaged.

3.3.1 Financial

The most well-understood incentive for performing microwork is the financial gains for completing tasks. Workers can use Mechanical Turk as a primary or secondary incoming, depending on their available time and interest, and may view the financial gains they receive as money to use for recreation, or as a safety net towards fulfilling their basic needs [Martin et al., 2014].

Martin et al. [2014] studied members of Turker Nation, a forum where turkers are able to discuss Mechanical Turk HITs and compare notes on specific tasks or requesters, in addition to socializing with one another. A major theme of discussion was the turker's self-perceptions of Mechanical Turk, which was primarily viewed as a labor marketplace where money was the primary benefit of using the program. Workers reported their strategies for hitting daily or annual financial targets,

and most desired for their work to recieve financial compensation and recognition comenserate with their level of effort.

3.3.2   Other Motivations

Initial investigations into motivations for participating on Mechanical Turk gave equal focus to financial and non-financial motivations, such as entertainment and killing time. However, especially on Mechanical Turk, workers may over or under-report altruistic motivations due to social desirability bias (where participants try to answer in the way they think is how they "should" answer, wanting to present themselves in a way that others will approve of) [Antin and Shaw, 2012]. While most research and self-reports now agree that turkers' motivation is primarily financial, looking at these other motivations can elucidate ways to design volunteer-based systems or to encourage higher-quality work.

Citizen science projects, or other tasks of a scientific or altruistic nature, may have natural intrinsic motivations. Participants want to contribute to the greater good, or help others, and see these projects as a way to do that. However, these tasks may also have other motivations, as participants may hope to gain social status or higher reputation among the community [Nov et al., 2011].

Non-financial motivations may impact qualities of the work that is received. Experiments have shown that in tasks coming from requesters who have a charitable bent (*e.g.,* non-profits), the work produced is higher quality than for corporations [Rogstadius et al., 2011]. In an experiment, workers were asked to mark circular objects in an image, with some workers told that the circular objects were cancerous cells, and others given no context [Chandler and Kapelner, 2013]. While there was no increase in speed, the workers given a scientific motivation stayed engaged with the tasks for longer (with no additional pay), indicating an impact of the task's motivation.

A personal relationship with a specific cause may also encourage workers or volunteers to participate. In a project creating guides to social interactions for people with autism, the researchers found that many workers from Mechanical Turk had been drawn to the task be-

cause they knew someone with autism or a similar developmental disorder [Hong et al., 2012]. These personal relationships can provide benefits both to system designers and workers - designers get more domain knowledge since workers have experience, while workers are able to contribute to a cause they care deeply about.

4

Examples of Human-Powered Access Technologies

The human crowds available via technology are clearly capable of performing a wide variety of tasks. These crowds can be leveraged for accessibility purposes, to help provide people with disabilities with information that would be hard to access otherwise.

In this section we describe 16 examples of human-powered access technology. These examples were chosen for the breadth of human support they represent, and will be referenced throughout the rest of the article.

ASL-STEM Forum

ASL users often do not know or cannot agree upon signs for terms in technical specialties, such as mathematics, science, or technology. This makes it difficult for Deaf students, researchers, or professionals to participate in the STEM field. The ASL-STEM Forum [Cavender et al., 2010] serves as a portal for ASL users to create and discuss signs for technical terms. Forum users can upload videos of signs for technical terms, watch and rank videos uploaded by others to learn more, or comment on videos with questions or suggestions. This model of a community-built forum

echoes that of Wikipedia [Viegas et al., 2007], where users can self-organize and contribute content that is important to them.

Tactile Graphics Project

Tactile graphics are the most useful format for presenting complex graphical images to blind users. However, they take time and expertise to create and are not readily available, so blind students are often at a great disadvantage when studying STEM topics. The Tactile Graphic Assistant (TGA) [Ladner et al., 2005, Jayant et al., 2007] aims to facilitate tactile graphic specialists by automating key parts of the process to reduce their workload. Text labels and descriptions can be automatically recognized and translated into Braille, and image features can be recognized and laid out in the correct format. While these steps can be automated, a human review of the graphics and a thorough editing process still needs to be completed by the specialists by hand – however, the automation components mean that less hands-on work is required. With the TGA, specialists are able to create a new graphic in an average of 10 minutes, a great improvement over the one to many hours previously required per graphic.

GoBraille

To enable safer and more independent public transit use for blind and deaf-blind people, GoBraille [Azenkot et al., 2011] allows users to access real-time bus information and physical details about bus stops (such as if they have benches or are covered) in Braille on their mobile phones. While the stop locations and real-time bus location information is sourced from publicly available web services, the physical details about the stops are sourced from the blind or dead-blind bus riders themselves, who can contribute information about stops they are using or rank others' descriptions. Rather than connecting GoBraille users with general transit patrons, the application allows sharing directly between disabled stakeholders, increasing the relevance of the information about the stops.

ESP Game and Games with a Purpose

Images on the web often remain inaccessible to blind or low-vision people, since developers often do not add alternative text [Lazar et al., 2004] and computer vision tools are not yet able to describe complex scenes. The ESP Game [Von Ahn and Dabbish, 2004] was created to enable sighted players to label images in a fun, low-effort, and social way. In the game, two remote players are shown the same image, and asked to guess a word that describes it. Once both players guess the same word, they are awarded points and the word can be used to label the image. Future players who encounter the same image may be told they cannot guess the previously assigned label, ensuring that multiple distinct labels can be collected for each image.

The ESP Game and other games with a purpose (such as Phetch, which generates longer image descriptions in sentence format [Von Ahn et al., 2006]) serve dual purposes - performing human computation tasks, like improving the accessibility of the web, while entertaining the players. Since the game format is less formal and tedious than traditional microwork, games with a purpose appeal to broader audiences, allowing more work to be completed without financial compensation.

Bookshare

While printed materials can easily be made accessible using OCR, there are copyright policies that prevent individuals from scanning and distributing media, preventing the creation of a free database of textual materials. Bookshare[1] is a website whose membership is open only to people with disabilities, exempting it from traditional copyright regulations[2]. The site aggregates books and transforms them into accessible formats, using the Digital Accessible Information System (DAISY) for talking books and screenreaders, or Braille Refreshable Format for refreshable Braille displays. In order to offer a large selection of books, volun-

---

[1]http://www.bookshare.org/
[2]Under US Code Title 17, section 121

teers participate by scanning books to contribute to the library, or proofreading the digitized books before they are offered to disabled readers. Through this method, Bookshare has made over 300,000 books accessible.

Respeaking

While automatic speech recognition (ASR) works well in ideal conditions, it cannot be used to accurately transcribe speech that is muffled or interspersed with background noises, and thus is not suitable for many real world scenarios. Audio streams can be cleansed for ASR through the process of "re-speaking" [Miyoshi et al., 2008], where a person repeats what they hear directly into a microphone, sometimes with the help of a speech silencer which covers their mouth. This practice of re-speaking audio can greatly improve the quality of ASR, since there is no interference with the text stream and the ASR can be trained for the specific respeaker in advance.

Social Accessibility

While many developers are not opposed to making their websites and products accessible, they often lack the time or expertise to do so [Lazar et al., 2004]. If a blind user encounters a web accessibility problem, they can request help from the developer directly, but are then required to wait until the developer finds time to implement changes, by which point the information may no longer be relevant. To improve the turnaround time of requests for help, the Social Accessibility Project [Takagi et al., 2008] lets blind users report problems online to sighted volunteers. These volunteers can provide metadata about the page (such as alternative text for images or structural information about text hierarchy), and that data is then available through a script that works with the users' screenreader. The collaborative and distributed nature of the Social Accessibility Project means that nearly half of requests can be filled within 24 hours, as there are many volunteers available to answer a call for help [Sato et al., 2010]. The data generated can also be useful for developers, providing them with

real instances of accessibility problems with their websites, and examples of how to fix them.

Remote Reading Systems

While automated systems can read clearly printed text well, they cannot easily read handwriting or extract knowledge from text. The Smith-Kettlewell Eye Institute introduced remote reading systems [Brabyn et al., 1992] in the mid-1990s, where blind people could fax documents to remote, sighted readers who would dictate the text over the phone. By connecting blind people and sighted readers across physical distance, remote reading systems facilitated independence and allowed blind people to access printed documents faster than if they had to wait for a reader to come to them.

Remote Real-Time Captioning

Real-time captioning translates streams of audio into text, utilizing stenography tools (where trained experts type a phonetic transcription of the speech, which is then converted to words) or computer-assisted notetaking (where speech is recorded as notes or verbatim text according to the typist's skill level) [Stinson et al., 1999]. This process can be performed remotely by streaming audio over phone lines or an internet connection, and is the standard way to caption telephone calls and television shows.

One of the most common forms of real-time captioning in schools is computer-assisted realtime transcription (CART), a form of stenography that can be performed in-place or remotely. Research has shown that access to a realtime stream of text provides significant benefits over sign language interpretation in learning outcomes [Marschark et al., 2006].

Scribe4Me

In scenarios without a hearing companion, Deaf people may be unaware that noise happened, or not be able to figure out what caused it. Scribe4Me [Matthews et al., 2006] is a mobile application that is constantly recording a 30-second 'buffer' of audio

for Deaf users. If they perceive that an important event has occurred, they can send the last 30 seconds of audio to a remote worker, and a transcription will be sent back to them within 3 to 5 minutes. Despite this delay, users found the service helpful for both capturing public information (*e.g.,* announcements over the intercom in an airport, or fragments of a conversation they were involved in that they had missed), and for gauging what was happening in conversations around them (*e.g.,* determining if a conversation among coworkers after a meeting was just conversational, or if it related to the meeting's topic and the user should participate).

MAPS-Lifeline

MAPS-Lifeline [Carmien et al., 2003] allows remote caretakers to create scripts for someone with a cognitive disability, creating a distributed support system that can enable the user to complete daily activities independently. The caretaker is able to monitor when the user starts an activity (as determined by sensors), and will be notified if the user goes off-track, ensuring that the caretaker is available to provide assistance and knows that the activities are taking place. This model of remote supervision is far cheaper and more efficient than having a caretaker physically present for the user.

Solona

CAPTCHAs are computerized tests used as gatekeepers to distinguish humans and computers and prevent automated attacks on websites, by asking them to perform a task (identifying obfuscated characters) that is easy for a human, but difficult for an electronic system. CAPTCHAs are a unique obstacle for blind users, as visual CAPTCHAs are inaccessible [Holman et al., 2007], but audio CAPTCHAs may be more difficult for the user and exclude users with disabilities [Bigham and Cavender, 2009].

Solona[3] was a service that let blind users send screenshots of CAPTCHAs to sighted volunteers, who tried to answer the

---

[3]http://www.solona.net/

CAPTCHA for them within 30 minutes. While this service is now shuttered, other web accessibility tools (such as Webvisum[4]) have stepped up to offer the same functionality, aiming to improve response times to be near-instant.

TapTapSee

TapTapSee[5] is an object recognition application for blind users. It recognizes objects from photographs, by using automated recognition when possible but falling back to crowd workers when automated tools are unsuccessful. The application has a simple interface and built-in camera corrections for blind users, but only performs object recognition and cannot answer more complex questions, and runs on a paid model (either paying by image sent, or subscribing for a month's worth of time.

Scribe

Scribe [Lasecki et al., 2012] is a crowd-powered system for real-time captioning. In this system, audio is streamed to multiple non-expert captionists, who each type part of the streaming audio that is played for them. The partial captions are then merged together to form a complete captioning output, meaning that even non-expert typists can create an accurate stream of text for Deaf users.

SocialMirror

SocialMirror [Hong et al., 2012] is a social network site that leverages the existing network of family, friends, and caretakers to provide support for adults with autism, helping them prepare for social situations and allowing them to ask for advice. Users can share their calendar with other members on the network, and ask questions about what to do for upcoming events or send pictures of themselves to their network to see if their clothing choices are appropriate. The application allows adults with autism to function independently, while still connecting them to trusted people who can help when needed.

---

[4]`http://www.webvisum.com`
[5]`http://www.taptapseeapp.com`

Similar social networking sites have been developed for other stakeholders, such as people with dementia [Martins et al., 2014]

5

Design Dimensions for Human-Powered Access Technology

The examples outlined in the last section collectively exemplify a large number of design dimensions that one might use to classify the a new human-powered access technology. These dimensions can be broadly grouped into three categories: *(i)* dimensions of the user of the technology, *(ii)* dimensions of the task being performed, and *(iii)* dimensions of the humans who power the other side of the technology. Here we describe these dimensions, refined from a list of design dimensions presented in [Bigham et al., 2011], and show an analysis of how the 15 example applications discussed above use these dimensions.

5.1 Dimensions of Users

The users for whom a human-powered access technology is designed have a significant impact on the tool itself. There are practical considerations based on the type and severity of disability of the user (*e.g.,* a tool for blind users must have a non-visual interface), and there are implications on the way the user interacts with the service.

5.1.1 Targeted Disability

Technologies vary based on what types of disabilities the tool is targeted towards. Sensory disabilities are widely regards as the most popular targets of automatic technology and this seems to be true for human-powered access technology as well. Applications for other types of disabilities exist, such as the MAP-Lifeline tool to support people with cognitive disabilities. We do not attempt to provide a breakdown of different types of disabilities, as many such classifications exist.

5.1.2 Initiative

Initiative refers to who instigates assistance: the end user, the workers themselves, organizations, or technology working on behalf of any of these parties. Individuals may be both users and workers, such as in Bookshare and GoBraille. Occasionally, initiative is mixed, such as when end users make requests and then workers choose which of the requests they will work on at a given time.

The *end user* often decides when to solicit help from human supporters. Examples include services like remote real-time captioning and relay services, and crowdsourcing systems like Social Accessibility and VizWiz. Systems where the end user initiates a request for help can serve dual purposes - in addition to giving assistance to the user in the moment, these tools can also be used retrospectively to learn more about what situations are inaccessible. Collecting data about the types of requests users have may help improve automated systems, or allow tools to appropriately scale for anticipated workload.

Systems like the ASL-STEM Forum and Go Braille allow *workers* to decide when and what information they will provide. For instance, what terms they will sign in the forum or what landmarks they will label in Go Braille. Worker-initiated systems are often those in which specialized expertise is necessary, since the smaller pool of available workers may necessitate a slower turnaround time for results and more flexibility in when the work is completed.

Organizations or groups of people will sometimes decide to solicit the help of human workers or to guide their efforts. In the ASL-STEM

Forum, workers are recruited to work on specific topic areas. These requests for help may create a wider and more diverse collective of workers than in other systems, as online contributions can be motivated by making people feel that their expertise is uniquely valuable [Ling et al., 2005].

Technology can be used as an intermediary to determine when assistance is needed for a person, or how to distribute tasks to get answers in a way that maximizes speed, quality, or throughput. The MAPS-Lifeline system uses sensors to determine if a cognitively disabled user has veered off-path. If so, their caretaker is notified and can check the system to determine if assistance is needed, meaning the caretaker does not need to perform constant monitoring of the user. Technologically-guided approaches can also result in systems with higher coverage than those where the workers decide themselves what to do: in the Scribe transcription model, the audio stream is split into small parts for workers to transcribe, making it more likely that all parts of the audio content will be transcribed (rather than allowing workers to choose what portions to transcribe, ending up with redundancies and holes in the transcription).

5.1.3   Confidentiality, Privacy, and Anonymity

When humans are included in the loop, new issues arise in terms of confidentiality, privacy, and anonymity. This dimension concerns how a human-powered access technology seeks to provide guarantees of safety or privacy to the users of the service, or how it makes trade-offs between this dimension and other concerns like latency or financing.

Trusted worker pools can reduce the perceived privacy concerns of human-powered access technologies. Remote interpreting services require workers to agree to strict confidentiality rules, and oversight helps to ensure that workers comply. Many professionally trained workers adhere to codes of conduct which provide these guarantees. Other tools may provide *feedback to users* to help them realize the implications of recruiting assistance. For instance, a tool might automatically detect that a face is included in a photograph before sending to an untrusted source of workers.

However, many tools provide *no guarantees* regarding confidentiality, privacy or anonymity. By making no guarantees, tools may have some benefits, *e.g.,* operating more cheaply, but users may not understand the risks they are taking by using the services. In VizWiz, some blind users inadvertently reveal private information in their photographs, whether it be purposefully revealing information but not understanding the risks (*e.g.,* sending pictures of credit cards and asking for the number), or not realizing that something has been captured in their photograph (*e.g.,* a piece of mail with the user's home address listed is sitting on the table next to the subject of their photograph).

5.1.4 Consideration of Broader Context

To the extent that privacy, anonymity and confidentiality are considered in human-powered access technology, the focus is generally exclusively on the end user. For instance, technology may go to great lengths to help protect the user's identity. Often, the effects on others in the broader context in which the technology is used are ignored. For instance, bystanders may unwittingly find themselves in the lens of a blind user using VizWiz. Workers may also be asked to answer questions with real-life consequences (*e.g.,* what is the dosage listed on this medication bottle?).

For the *user's consideration*, mechanisms should be in place to match the user's expectations. Video Relay Services (VRSs) do this well - the services are initiated by the user, with a clear indication when the user is being recorded, and VRSs are regulated by the FCC. For *worker's consideration*, workers should be allowed to refrain from providing assistance in situations in which they feel uncomfortable. Workers may be told the consequences of their assistance and allowed not to assist without penalty.

Mechanisms may also be in place for *community consideration*, to protect people who are not directly involved in the use of the technology from accidentally becoming part of an interaction. Services like Scribe4Me may inadvertently violate the privacy of others by recording their conversations and passing it to an external transcriptionist without their knowledge.

5.2 Dimensions of Tasks

The type of need the service aims to meet also may impact how a human powered access technology should be designed. Tasks may be short and asynchronous, or longer-form and requiring a paid worker to remain available for a longer period of time.

5.2.1 Financing

Depending on the source of human workers and their expertise, different funding mechanisms can be utilized. The type of financing that supports a human access technology may have implications for the quality, availability, or sustainability of the service.

Many human-powered access technologies are *publicly funded*. For instance, video relay services are usually publicly funded, subsidized by fees that everyone pays on their phone bill. Other services, like closed captioning for television, may be supported by a combination of government or advertiser funds. In other cases, human assistance is *funded by the user*, either personally or through the user's workplace, educational institution, or insurance. Private funds and trusts fall in the same category. Even semi-automated tools, like TapTapSee, incur ongoing expenses. TapTapSee allows users to buy "packs" of questions, or to subscribe to the service for month-long periods of unrestricted question asking.

Some assistance is provided *for free*. For instance, the ESP Game attracts workers through a game. Both Solona and Social Accessibility were operated by volunteers and therefore free to use. To recoup operating expenses, though, services like this often have external funding sources - research projects like Social Accessibility and VizWiz are funded by grants from corporations and universities.

5.2.2 Latency

Different human-powered access technologies may have different expected latencies, based on the source of human computation and what the technology does to help achieve faster responses. It might be acceptable to wait a few hours (or even days) for a volunteer to read your

mail, but sign language interpreting may need to happen right away to enable conversation.

Some human-powered access technologies are *interactive*, allowing the user to get information near-immediately from the system. For instance, video relay services work in real-time. The Scribe system for real-time captioning has to work within the constraint of 5 seconds to allow students to pair the captions with the rest of the class context.

Many services operate under the assumption of *a short delay*. For instance, VizWiz and TapTapSee aim to return answers with minimal delay, in less than 90 seconds. Social Accessibility takes on the order of a few hours to resolve an accessibility problem on average, though some queries may languish for days depending on volunteer availability. In cases where the answer takes a long time to arrive, the end user is likely gone or engaged in other activities before the content they would need is available to them.

Some services have an *undetermined delay*. For instance, workers may never choose to add a video for a particular term on the ASL-STEM Forum, or they may add it immediately. Undetermined delay services are frequently those where the action is initiated by the worker, instead of the user requesting assistance.

5.2.3   Accuracy Guarantees

Humans can provide services that are still too difficult to automate. However, relying on humans may result in lower or unpredictable quality of work or assistance for a number of reasons. Workers often misunderstand what to do due to underspecified tasks, and can answer incorrectly due to fatigue or a rush to move on to their next assignment. Human-powered access technologies use a number of methods for quality control.

Many human-powered access technologies improve accuracy by using only *vetted workers*. Workers may be vetted by their employer, by volunteer organizations, or via reputations gained over time. This is how the Video Relay Service and Remote Reading Services work (among others). These services can be costly and may need to be arranged in advance.

Some technologies use *abstractions* to help guarantee accuracy. For instance, the abstraction introduced by the ESP Game (single word or phrase descriptions) makes it unlikely for the two game players to agree on a label if their inputs do not accurately describe the image. One way to help ensure accuracy is by requiring *redundancy*, where multiple workers must agree on the same answer. For instance, the ESP Game requires that multiple pairs of players submit the same label before it is counted as good enough to be passed onto the user.

In some cases, accuracy is *mediated by the user*, where users are presented with alternatives provided by human assistants and asked to decide whether they are correct. For instance, VizWiz provides users with answers from multiple workers and allows them to draw their own conclusions about their value, while GoBraille asks users to rank the user-submitted information they receive about bus stop landmarks while they are at the stop, allowing the user to verify the information in-person. Humans sometimes serve not only as workers but also as *human mediators*. In the ASL-STEM Forum, contributors help to decide which signs work best by contributing new signs and rating the ones that are already there. The crowdsourcing literature proposes many workflows for doing human verification of other human work, either allowing them to review and verify their own solutions [Dow et al., 2012] or calling in additional workers to check the work that has been done [Bernstein et al., 2010].

5.2.4 Reliability

Many sources of human assistance are not always available; for instance, a simple approach to engage a human worker would be to simply send a text message to a friend or family member. This method is unreliable because the friend or family member may not always be available. Many of the methods of recruiting human assistance assume a network connection, which may not always be available (*e.g.,* in a basement conference room or on a plane during takeoff). Traditional approaches, like hiring a sign language interpreter, do not face these reliability concerns.

Few sources of human assistance are *always available* as networking connectivity is a requirement, if nothing else. Nevertheless, services like

VRS employ workers who are available at all times, increasing reliability. VRS guidelines dictate that a call must be answered within two minutes and thirty seconds, and if a user tries to utilize the telecommunication relay services but cannot reach an operator, they are able to complain directly to the FCC.

Services built on microtask marketplaces (like VizWiz) are *assumed to be reliably available*, but there are not guarantees that this will always be the case. As discussed above, the Solona service had initially been available nearly 24/7, but had to scale back since its' creator was the main CAPTCHA answerer and could not sustain full hours by himself.

Similar to the latency dimension, other services like the ASL-STEM Forum have *undetermined reliability*. While people may come to contribute and rate signs, they may come infrequently or stop using the service without warning. Tools that rely on a "critical mass" of users may no longer be able to function if people stop using them, *e.g.,* the ESP Game will not work without at least 2 people playing.

5.2.5 Broader Applicability of Human Work

The work that humans perform in the context of access technology can be more broadly applicable to other people. In some cases, *individual reuse* may be possible for the work done on a user's behalf. VizWiz allows users to store the descriptions of their photographs in the application, so if they want to revisit a description, they can go back to it easily. Many services allow *group reuse* the work that was done on the behalf of others. For instance, metadata improvements for web pages in Social Accessibility can be used by other people who visit the same web page. The annotations made in Go Braille can be used by other travelers, and books that users scan for personal use can be added to BookShare for others to access.

In other cases, there can be *no reuse*, either because it does not make sense to reuse work or because the application doesn't support it. For instance, in VRS it does not make sense for others to reuse the sign language interpretation of a conversation. While some work cannot be directly reused, they can contribute to *scientific reuse*. The results of the initial use can be used later to build smarter automated tools. The

answers to VizWiz could be used to train better object recognition tools, or to learn about common photographic errors that occur for users who are blind and build tools to correct for them.

5.3 Dimensions of Workers

Workers can come from different sources, which may influence their motivations for performing work, their level of skill or ability, and how the work is attributed to others.

5.3.1 Source of Human Workers

A number of sources of human workers now exist, from professional organizations and microtask marketplaces to friends and family on social networks. These sources have a variety of different characteristics (*e.g.,* speed, quality, availability) which may necessitate tradeoffs between them, and could also be combined to form hybrid systems.

Professionals are hired to provide assistance and generally receive wages for their participation. Stenographers are professionals who must be trained in phonetic typing before they are able to use a phonetic keyboard, and as such their services can be expensive. A number of business models may support these workers - in some cases, they are hired by specific organizations to make their services more accessible (*e.g.,* a company hiring a stenographer to caption an event); while in other cases the costs are subsidized by the government (*e.g.,* TTY systems for phone calls) or the beneficiary must pay for services themselves.

Crowd workers are recruited anonymously through an open call, are generally (but not always) composed of non-experts, and cannot be assumed to stay around for a significant amount of time. Platforms like Amazon's Mechanical Turk allow workers to sort through tasks and select those that interest them or have an appealing price point. Other platforms, like the non-profit Samasource[1], distribute generic work directly to workers to complete and pay a fixed rate.

Volunteers work for free, are generally not anonymous, and generally participate for a longer periods of time than crowd workers. Rather

---

[1]`http://www.samasource.org`

than working for payment, volunteers participate for more altruistic reasons - to help people, contribute to scientific good, or because they had a pre-existing interest in the cause their work will benefit [Raddick et al., 2010]. One platform for online volunteering is Help from Home[2], a UK-based microvolunteering platform where volunteers can search for volunteer tasks by the expertise they require, the cause they will help, and the amount of time they will need to commit. Volunteer-run projects may be hard to sustain over time. As mentioned above, Solona was sustained primarily by it's creator, Bernard Maldonado, who eventually had to cut back its' operating hours to ensure that he was able to get a full nights' sleep [Solona, b].

In addition to unstructured volunteer platforms like Help from Home, *organized volunteers* can also be recruited through an group or organization. Organizational oversight may help ensure that people are available when needed, that worker skills are vetted before they start with tasks, or that structured training is provided.

5.3.2 Motivation

As in everything we do in our lives, human workers have a variety of reasons for helping others. These reasons range from altruistic intrinsic motivation to the desire to make money.

People are often willing to contribute their time for *intrinsic reasons*, like causes they deem valuable. They may be motivated by the fact that their relatives or friends have disabilities, or just want to help without expecting anything in return. SocialMirror and other friend-sourced access technologies harness this intrinsic motivation by connecting people with disabilities directly to their friends and family members as a resource. Even in projects where the worker has no direct relationship with the beneficiary, they may work longer if they believe their work is benefitting a good cause [Chandler and Kapelner, 2013].

Others are better motivated when their contributions contribute to their *status*, by being made public through announcement, awards, and banquets. These people can often be encouraged to work by offering gains to their public status, or tasking them with completing a specific

---

[2]http://www.helpfromhome.org

objective or achievement. The Bookshare website highlights profiles of notable volunteers on their website, which may make people want to do good work to be known in the community. In the ESP game, a scoreboard promotes competition between players and motivates people to contribute more captions.

Some people are best motivated by *financial rewards*, either in small amounts for fulfilling tasks at microtask marketplaces or as adequate compensation for their professional services. Crowdsourcing platforms typically are driven by financial exchange, though the impact of financial rewards may be more significant for workers in developing countries like India than in the United States [Ipeirotis, 2010].

Some systems make it *entertaining* to provide assistance. The ESP Game and Phetch both turn image labeling into a game, and many workers on Mechanical Turk report participating in tasks for fun, or working longer on tasks that are enjoyable than those that are boring [Martin et al., 2014]. Users may also provide assistance because it *helps themselves*. For instance, users may contribute books that they scanned for themselves to Bookshare for shared use. One proposed approach for the Scribe system, where workers collectively caption a live audio stream, is that it can be self-motivated for educational situations by having students work as the captioners during their class. While the Deaf student gets live captioning, all students then have access to a transcription of the class afterwards, which they can use while studying.

5.3.3 Worker Competence

Different sources of human work may provide workers with different competencies. Even within a particular source, worker competence may vary dramatically. Human-powered technologies use a variety of methods, from workflows to statistical post-processing, to improve the experience quality of work. For people with disabilities, the output of access technologies can be crucial for conducting daily activities independently, so worker competence must be assessed and accounted for when building tools.

Many services will require *skilled workers*. For instance, a video relay service requires proficient sign language speakers, and the Tactile

Graphics Project can only help trained graphic makers by shortening their initial workload, but cannot replace their expertise in doing the final alignment of figures and labels in the tactile representations.

Some services may be supported by *amateurs*. For instance, the ASL-STEM Forum invites anyone to contribute, under the assumption that the rest of the community will filter out the good contributions from the bad. For Bookshare, members need to have access to some type of OCR tool to scan their books, and must become familiar with the DAISY and BRF formats before proofreading scanned texts.

Many crowd-powered systems are designed so that assistance can be provided by *non-experts*. For instance, VizWiz uses workers recruited from Mechanical Turk. In this case, the ability to see is the only qualification. Other systems, such as Scribe, combine the effort of multiple non-expert workers to approach the performance of an expert stenographer. While these tools do not require any special skills or expertise, the quality of the work is still important, and must be verified or backed up.

5.3.4 Assistance Provenance

Providing feedback to users about the human computation that is occurring on their behalf is critical for them to make informed decisions. In particular, it can be useful to know who or what has provided assistance. Automated tools may fail in certain cases, or provide misleading information; the quality of human work may vary based on if the worker is an expert or an non-expert.

In *transparent* systems, users are informed regarding who or what will be assisting them beforehand. Most VRS operate in this way, where users are first connected to the person that will be assisting them. *Obfuscated* systems tell users the general source or sources of assistance, but they have no way of knowing who or what actually assisted them for a particular request. TapTapSee works this way - pictures are described by either computer vision or human workers, but are not labeled as to the source of assistance.

In *opaque* systems, users are not informed who or what assisted them. The service operates like a "black box", which may lead to con-

fusion for the end user. TapTapSee, which relies on a combination of automatic and crowdsourced object recognition, does not identify what source each answer is from, so users have no way to know if a human has looked at their photograph or not.

5.4   Application of Design Dimensions

These design dimensions can be used to analyze and study human-powered access technologies. In Figures 5.1, 5.2 and 5.3, we show how the 15 applications from Section 4 applied these dimensions.

	ASL-STEM Forum	Tactile Graphics Project	GoBraille	ESP Game	Bookshare
Initiative end (U)ser, (W)orkers, (O)rganizations	W	W,O	U,W	W,O	U,W
Workers (E)xperts, (C)rowd, (V)olunteers, (O)rganization, (A)utomation	C,V,O	P,V	C,V	C	E,V,O
Motivation (I)ntrinsic, (S)tatus, (F)inancial, (E)njoyment, (P)ersonal	I,S	I,E	I, P	S,E	I,F,P
Financing (P)ublic, pe(R)sonal, (U)npaid	U	P,U	U	U	P,R,U
Competence (E)xpert, (A)matuers, (N)on-expert	E,A	E,A	A, N	N	E,A
Latency (I)nteractive, (S)hort delay, (U)ndetermined	U	U	U	U	U
Accuracy (V)etted, (R)edundancy, (U)ser mediated, (W)orker mediated	U,W	V	U	R	V,U,W
Reliability (A)lways, a(S)sumed, (U)ndetermined	U	U	U	S	S
Assistance Provenance (T)ransparent, (O)paque, o(B)fuscated	T	B	B	B	B
Confidentiality (T)rusted workers, (U)ser feedback, (N)one	N	N	N	N	N
Context (U)ser, (W)orker, (C)ommunity, (N)one	U	N	U, W, C	N	U,W,C
Application (I)ndividual, (G)roup, (N)one	G	G	G	G	G
Target Group	Deaf	Blind	Blind	Blind	Print

Figure 5.1: An analysis of the 15 example applications on the 13 design dimensions for human-powered access technology (Page 1 of 3). Reported values are simplifications but support analysis of design trade-offs and comparison. For instance, VizWiz and Solona are both human- powered systems for describing images, but they use different sources of human labor. Solona's expert workers increased confidentiality, competence and accuracy, but at the cost of latency and reliability. The letters used to indicate each value are identified in the header row with parenthesis.

	Respeaking	Social Accessibility	Remote Reading Services	Remote Realtime Captioning	Scribe4Me
Initiative end (U)ser, (W)orkers, (O)rganizations	U,O	U	U	U	U
Workers (E)xperts, (C)rowd, (V)olunteers, (O)rganization, (A)utomation	E,V,O	C	E,O	E	E
Motivation (I)ntrinsic, (S)tatus, (F)inancial, (E)njoyment, (P)ersonal	I,F	I,S	F	F	I
Financing (P)ublic, pe(R)sonal, (U)npaid	P,R,U	U	P,R,U	P,R	U
Competence (E)xpert, (A)matuers, (N)on-expert	E,A	A,N	E	E	A,U
Latency (I)nteractive, (S)hort delay, (U)ndetermined	U	S	I	I	S
Accuracy (V)etted, (R)edundancy, (U)ser mediated, (W)orker mediated	V,W	U	V	V	V
Reliability (A)lways, a(S)sumed, (U)ndetermined	U	S	A	A	A
Assistance Provenance (T)ransparent, (O)paque, o(B)fuscated	T	B	T	T	B
Confidentiality (T)rusted workers, (U)ser feedback, (N)one	T	N	T	T	T
Context (U)ser, (W)orker, (C)ommunity, (N)one	U,W	U	U,W	U,W	U
Application (I)ndividual, (G)roup, (N)one	N	G	N	N	U
Target Group	Deaf	Blind, Other	Blind	Deaf	Deaf

Figure 5.2: An analysis of the 15 example applications on the 13 design dimensions for human-powered access technology (Page 2 of 3). Reported values are simplifications but support analysis of design trade-offs and comparison. For instance, VizWiz and Solona are both human- powered systems for describing images, but they use different sources of human labor. Solona's expert workers increased confidentiality, competence and accuracy, but at the cost of latency and reliability. The letters used to indicate each value are identified in the header row with parenthesis.

	MAPS-Lifeline	Solona	TapTapSee	Scribe	SocialMirror
Initiative end (U)ser, (W)orkers, (O)rganizations	W	U	U	U	O
Workers (E)xperts, (C)rowd, (V)olunteers, (O)rganization, (A)utomation	E	E,O	C,AI	C,V,O	C,V
Motivation (I)ntrinsic, (S)tatus, (F)inancial, (E)njoyment, (P)ersonal	F	I	F	F,E	I,F,P
Financing (P)ublic, pe(R)sonal, (U)npaid	P,R	U	R,U	P,R,U	P,R
Competence (E)xpert, (A)matuers, (N)on-expert	E	A	N	A,N	A,N
Latency (I)nteractive, (S)hort delay, (U)ndetermined	S	S	S	I	U
Accuracy (V)etted, (R)edundancy, (U)ser mediated, (W)orker mediated	V	V	V	W	V
Reliability (A)lways, a(S)sumed, (U)ndetermined	A	S	A	U	U
Assistance Provenance (T)ransparent, (O)paque, o(B)fuscated	B	B	O	O	B
Confidentiality (T)rusted workers, (U)ser feedback, (N)one	T	T	N	N	T
Context (U)ser, (W)orker, (C)ommunity, (N)one	U,W	U	N	U,C	U
Application (I)ndividual, (G)roup, (N)one	N	N	G	I,G	I,G
Target Group	Cognitive	Blind	Blind	Deaf	Developmental

Figure 5.3: An analysis of the 15 example applications on the 13 design dimensions for human-powered access technology (Page 3 of 3). Reported values are simplifications but support analysis of design trade-offs and comparison. For instance, VizWiz and Solona are both human- powered systems for describing images, but they use different sources of human labor. Solona's expert workers increased confidentiality, competence and accuracy, but at the cost of latency and reliability. The letters used to indicate each value are identified in the header row with parenthesis.

6

Case Study 1: Making Web Pages Accessible with Social Accessibility

The Web holds great promise for people with disabilities. Unlike printed material that can be difficult for some people to use and difficult to change, content on the Web is represented in electronic formats whose presentation can be adapted to fit the needs and abilities of diverse users. For example, screen reading programs convert text to voice or refreshable Braille so that a blind person can read them, and magnifiers can increase the size of text so that people with low vision can effectively read it [Bigham, 2014, Kline and Glinert, 1995]. Unfortunately, to realize these benefits, the web page must be created with appropriate metadata to allow an access technology to understand it.

One stubbornly persistent example where existing web content is lacking is in web images. Web images are inaccessible to blind users browsing the Web with a screen reader *unless* the author of the web page thought ahead to assign the image a textual description, *i.e.,* an alternative text description. Despite being one of the most obvious examples of an accessibility problem, only 20% to 40% of images on the Web have alternative text descriptions provided (and many of these are unhelpful, *e.g.,* "image", "button", "" [Bigham, 2007]). The proportion of images with descriptions has changed little over the last 10 years, even

among government web sites that are compelled to produce accessible content [Olalere and Lazar, 2011].

Other metadata can make other content easier or possible to use. For instance, structural tags like headings allow for more efficient navigation through web content. While content creators can provide metadata for their own pages, many images online still lack alternative text [Hackett et al., 2004], and missing alt-tags and layout information are among the most frustrating problems encountered by screen reader users [Lazar et al., 2007]. While developers may support the concept of accessibility, they often lack the knowledge or time needed to make their sites accessible [Lazar et al., 2004], and if a user relies on the site's owners to fix accessibility problems, it may take so long to contact them and get a correction that the information is no longer needed. This problem increases over time as we amass a legacy of web content that was not created in an accessible way and which may no longer be actively maintained.

Social Accessibility is a human-powered access technology that took the different approach of allowing third-party volunteers to fix problems on the web pages of others [Takagi et al., 2008]. While prior projects had explored this idea, *e.g.,* accessibility-oriented scripts in Greasemonkey [Pilgrim, 2005], Accessmonkey [Bigham and Ladner, 2007], and AxsJAX [Chen and Raman, 2008], Social Accessibility was rare in the breadth of problems it allowed volunteers to address and unique in the scale at which it operated.

In this chapter, we discuss the design of Social Accessibility, the use of the system in practice, and future research and lessons that can be drawn from the use of this system in practice.

6.1 Social Accessibility

Traditional metadata generation has fallen entirely on the shoulders of site creators, who must identify structural or layout information and images which need descriptions, generate that information, and add them as metadata on the webpage. Social Accessibility [Takagi et al., 2008] expands this ecosystem to include *(i)* screen readers users, who

identify problematic web content, and *(ii)* other people online, who can volunteer to generate the metadata based on the users' request.

When users of Social Accessibility are browsing the web and encounter an image without alternative text or a section of the page that lacks metadata, they can activate a JAWS Script which prompts them to describe the problem they have encountered, identifies the location on the page they are currently focused on, and sends this information (along with a screenshot of their current browser window) to the Social Accessibility server. Volunteers, who have installed a browser extension when registering with Social Accessibility, receive a notification that a new request has been made. If they are available to help, they can open the extension and are shown a rendering of how the screen reader reads the webpage; they can then insert metadata to resolve the user's request.

Volunteers are able to request clarifications from the users, or ask other volunteers for assistance, and an email is sent to the user once their request has been filled, notifying them that they can now return to the page and the new metadata will be inserted by the JAWS Script. Any Social Accessibility user can then access this metadata when browsing the page, and the user who initiated the request can provide feedback about the metadata provided, or request follow-up information. Based on how positive the feedback from the user is, volunteers are awarded points which are displayed on their personal profiles, indicating how many accessibility fixes they have contributed to.

6.2   Design Dimensions of Social Accessibility

The design dimensions presented in Chapter 5 can be applied to Social Accessibility, for a better understanding of how it fits into the landscape of human-powered access tools.

6.2.1   Users

Social Accessibility was targeted specifically for screen reader users, as the original version was developed for the JAWS screen reader. However, Social Accessibility could also be useful to people with other dis-

abilities. The structural information (*e.g.,* heading tags) could be useful to people with motor impairments, who may navigate with keyboard or voice commands instead of the mouse.

Metadata requests in Social Accessibility are initiated by the end user as they navigate a webpage and encounter something inaccessible or confusing. These on-demand requests mean that volunteers are not aiming to make the whole web accessible (an impossible task, given the always-changing and expanding nature of the internet), but are able to respond to individual needs as they arise.

The original version of Social Accessibility sent screenshots of the blind users' computer to the volunteers, creating a risk of revealing private information that was also on the screen. In later versions, volunteers instead added metadata to a page map of the website, reducing the risk of showing other windows on the users' screen.

6.2.2 Tasks

The type of tasks performed in Social Accessibility may vary in type and length, though they are all related to metadata authoring. Volunteers may be asked to provide descriptions for images (a relatively quick, one-off task), or may need to tag the structure of an entire page to make it more logical for the reader (a longer task). Volunteers can also contact the user for clarification or more detail, which may delay the completion of the task.

The volunteer-only nature of Social Accessibility means that the availability of people who can answer questions is not guaranteed at any given time; some requests took over a week to answer, while other requests were never answered at all. These volunteers must also be considered as non-experts, which lead to the creation of an improved interface for tagging to makes it easier for people unfamiliar with metadata to participate. Users are asked to give feedback about the results to their requests, so if a request is not answered sufficiently more volunteers can be called in or the task can be passed to an expert.

Since metadata is retained after it is generated and shown to future visitors to a site, it has both individual and group reusability. The metadata can also be used by web developers to improve their own

sites, or by researchers who want to identify problems and improve web accessibility.

6.2.3   Workers

The volunteers for Social Accessibility are not organized as a group, but instead are unaffiliated with each other. While volunteers can converse with each other through the collaborative features of the platform, much of their work is done independently, and their identities are obfuscated and interchangeable to the users.

Volunteers are not paid for their involvement, but are motivated by at least two non-financial factors - status, which comes from the point-based measurement of each volunteers' contributions, and an intrinsic motivation from helping users and getting feedback that their assistance was useful. While little expertise was necessary to author metadata with the page-map tool, many users seemed to know about basic web accessibility, and may have contributed due to a desire to apply their knowledge in practice.

6.3   Evaluation of Social Accessibility

Following the technical development of the metadata and authoring tools for Social Accessibility and a small pilot test [Takagi et al., 2008], it was more widely released to examine how the system would work in practice over ten months [Takagi et al., 2009]. This released version included two major improvements on the original pilot: first, expanding the tool to be screen-reader independent, rather than using JAWS Scripts; second, improving the metadata authoring tools through a page map which identified untagged areas and a quick-fix tool which allowed iteration through each image on the page without alternative text.

Similar to web accessibility problems identified in prior work [Lazar et al., 2007], the majority of the 323 requests made by users of Social Accessibility were for missing alternative text (38.4%) or a lack of structural heading tags (24.8%). 275 (85%) of these requests were resolved during the study period, but since requests could be broad (*e.g.,* "Request for heading tags"), this resulted in a large amount of generated

metadata, with 11,969 images tagged with alternative text and 6,069 heading tags added.

Some requests could not be answered by the volunteers, who could identify requests that should be forwarded to an accessibility expert with more experience, or indicate that the request could not be solved via Social Accessibility.

The release of Social Accessibility revealed that this approach could allow volunteers, who may have had no formal training in accessibility, to improve the accessibility of webpages created by others on-demand. This model may not resolve all problems that users encounter, as users may not know that part of a page is inaccessible to them and some of the requests made by users went unanswered or needed expert input. However, the majority of requests made were simple (for alternative text or H1 through H6 heading tags), were answered within a week, and provided information for the requesting users and for future visitors to the sites.

6.4 Conclusions

Social Accessibility provided a valuable examination of how remote volunteers could contribute to web accessibility, and contributed an understanding of the types of accessibility problems that screen reader users encounter. While there are no guarantees that web developers will make pages accessible, this model of spreading the workload among a large group of distributed volunteers can make sites accessible faster than tracking down developer contact information and waiting for them to propagate changes.

The ability to request expert review of tough accessibility problems, combined with the design of the page-map tool for non-expert annotators, meant that even volunteers without accessibility training could feel comfortable contributing. This type of tiered reliance on workers, where non-experts could reach out for help or tag the problem for experts; more experienced volunteers could make complex changes or give feedback, means that the pool of volunteers can remain large rather than restricting participation to trained accessibility consultants.

The data from the pilot study revealed a number of problems with alternative text for images and heading tags throughout the page. However, the authors note that the system is limited by the users' perception of accessibility problems. If part of the page is inaccessible to the user, but they have no indication that it exists, they may never realize that content or functionality is unreachable and will not request help. The application of web accessibility standards, or implemention of site-wide accessibility fixes, could correct for this potential limitation.

While Social Accessibility was able to incorporate metadata immediately for new visitors after it was generated, the latency of the first request was significant, often measured in days. This means that the requesting user could at best benefit from their request on a subsequent visit, and likely had to satisfy their original information need in another way.

The next chapter covers VizWiz (Chapter 7), which is in some ways similar to Social Accessibility, but which targeted very low-latency answers to visual questions about the world around a user, which likely could not be repurposed by other users.

7

Case Study 2: Answering Visual Questions Quickly with VizWiz

For people who are blind, visual information in the physical world is often inaccessible. While some types of content are made available as the result of legal requirements (*e.g.,* room numbers in Braille), and other types of content can sometimes be accessed using automated tools (*e.g.,* text available through OCR tools), most visual information in the world is difficult to learn about, and requires the assistance of a sighted person to access.

In this chapter, we discuss *VizWiz*, a mobile phone application for blind users that connects them to sighted, human answerers to get answers to visual questions within a few seconds. We describe in detail the system design and its use. A unique aspect of VizWiz is the speed at which it leverages human workers to solve accessibility problems.

7.1 VizWiz

VizWiz is an accessible mobile application that allows users to take a photo, speak a question about it, and get answers quickly from the crowd [Bigham et al., 2010]. VizWiz does not attempt to be a full sensory substitution system, as converting all of what is captured in a
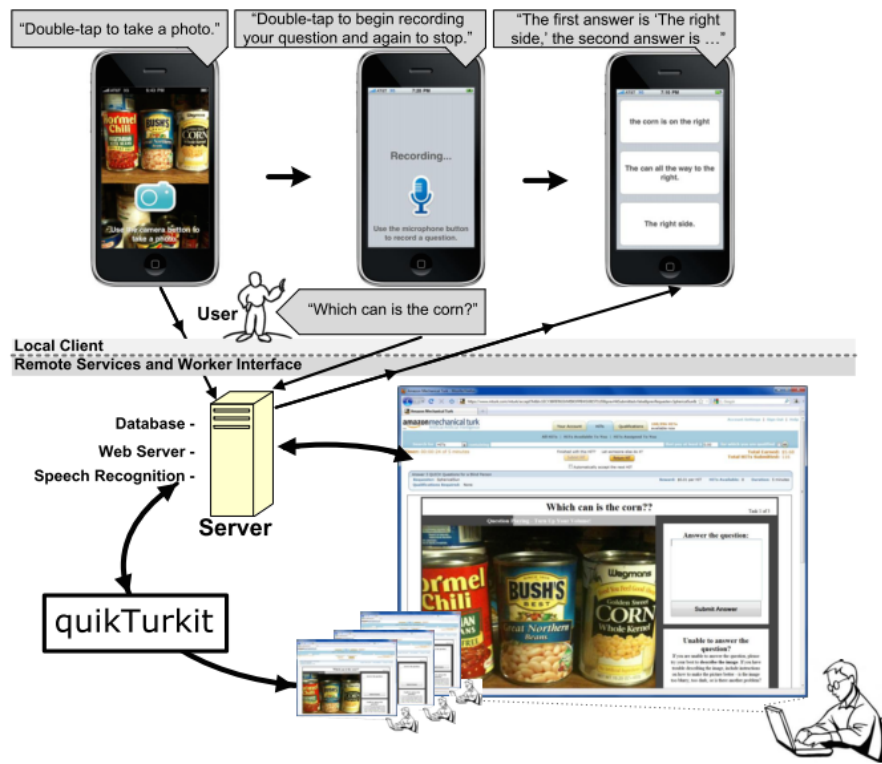
Figure 7.1: Original system diagram for VizWiz. A user can take a photograph, record their question, and send it to a worker. On the backend, a worker is recruited (when the camera is opened), the photograph is sent to them along with the audio and a transcription, and answers are returned to the user. From Bigham et al. [2010]

photo to speech is intractable. Instead, it relies on the blind users of VizWiz to scope the information needed by asking a question. Even as computer vision systems are able to describe some aspects of an image, they will not replace the functionality of VizWiz until they are able to scope the information provided to the needs of the user.

The initial investigation of VizWiz contributed *(i)* the design of the system, *(ii)* quikTurkit, a method for pre-recruiting workers to ensure fast responses to questions, and *(iii)* a field study of the application. These contributions, as well as lessons from the public deployment of the application, are discussed in Sections 7.3 and 7.4.

7.2 Design Dimensions of VizWiz

The design dimensions presented in Chapter 5 can be applied to VizWiz, for a better understanding of how it fits into the landscape of human-powered access tools.

7.2.1 Users

VizWiz is targeted towards people who are blind or have low-vision. Requests are entirely initiated by the user when they encounter some visual information in their environment which they cannot access.

The privacy implications and broader considerations of the interactions that VizWiz facilitates are serious. Users can easily take pictures of sensitive information, either inadvertently because they cannot review their photographs, or because the answer is important to them and they are willing to take the risk to get a prompt answer. This risk extends to the workers, as well, since photographs are not screened and so the worker may see inappropriate materials.

7.2.2 Tasks

VizWiz is currently supported by research grants, which pay the cost of the Mechanical Turk workers. Due to quikTurkit, answers can be recruited in nearly-realtime, and the large scale of workers on Mechanical Turk mean the system is assumed to be always available. Redundant answers are collected to each question, but it is left to the user to decide the quality of each response. While individuals can reuse the descriptions of their own images, there is little potential for group reuse, as the questions are specific to the user's environment.

7.2.3 Workers

Workers are primarily recruited from Mechanical Turk, who are motivated by financial incentives. However, as prior research shows, workers may be willing to work for longer periods of time if a task is related to altruistic cause, due to intrinsic motivation [Chandler and Kapelner, 2013]. The tasks require little expertise beyond the ability to see, so

non-expert workers are sufficient. The assistance provenance is obfuscated to the user, who knows a person is answering their question, but doesn't know anything about the worker.

Friends and family members can also be answerers for VizWiz questions. These workers are motivated by intrinsic motivations, and the assistance provenance in these cases is transparent to the user. It might be reasonable to think that questions could be answered by leveraging social sources like Facebook – however, a study found that VizWiz users preferred not to send their questions to their (extended) social networks due to concerns about bothering their friends or being perceived as dependent [Brady et al., 2013b].

7.3 Interface and Implementation

Here, we discuss the design of the interface for VizWiz, and details about the system implementation that supports it.

7.3.1 User Interface

VizWiz was built on the iOS platform in order to use the built-in VoiceOver screenreader. The system was designed for voice input (the user speaking into the phone) and text-to-speech output (from VoiceOver). While questions could be entered manually, the typing interface for iOS with VoiceOver enabled can be tedious, and speech recognition (*e.g.,* Siri) had not yet been built into the operating system. This interface also imitates other access technologies, like the kNF-BReader, which performs OCR on the text recognized in a scene and reads it aloud.

The user interface for the original VizWiz system is shown in the top half of Figure 7.1. When a user opened the application, it defaulted to the camera and VoiceOver read, "Double-tap to take a photo". Users could point the camera to an object or setting of interest, and snap a photograph by double-tapping anywhere on the phone screen. Then the user was prompted to double-tap to begin recording their question, and double-tap again to end the recording. Then, the user was taken to an answer screen, which instructed them to wait until an answer

arrived. Answers were automatically announced by VoiceOver when they arrived, or users could browse through them manually as desired.

For the original version of VizWiz, no computer assistance was available for the user when taking their photograph. Due to errors encountered in the first field study, consistent with other research indicating that photography is difficult for blind people [Vázquez and Steinfeld, 2012, Jayant et al., 2011], the next iteration included tools to detect lighting levels and blurriness in the photographs. If a photograph was too dark or blurry, the user was notified after taking it and given the option to proceed or retake it.

7.3.2 Implementation

On the backend of the system, a server waits for notification from the VizWiz application (as seen in the lower half of Figure 7.1). When the user opens their application, the server sends a notification to the server, indicating that a question may be forthcoming. Depending on the worker recruitment strategy (discussed further in Section 7.3.3), the server may begin recruiting workers at this point if none are currently available. Once the user takes their photograph and sends their question in, the server posts it to workers on Mechanical Turk. The worker interface, seen in Figure 7.2, presents a large image to the worker, and automatically plays their recorded question for the worker.

The worker is instructed to answer the question to the best of their ability, but if they are unable to answer due to photographic errors, describe the problem with the photographs so that the users can retake it and resend their question. Answers are collected from multiple workers, and forwarded back to the user as they arrive.

In the initial field studies of VizWiz, the workers also could see a transcription of the users' question, generated by automated speech recognition software. This transcription was often highly unreliable (as described in Section 7.4.1), but could be useful when transcribed well.

7.3.3 Crowd Workers

In order to ensure that human-powered answers are available at any time the user may have a question, VizWiz provides an option for
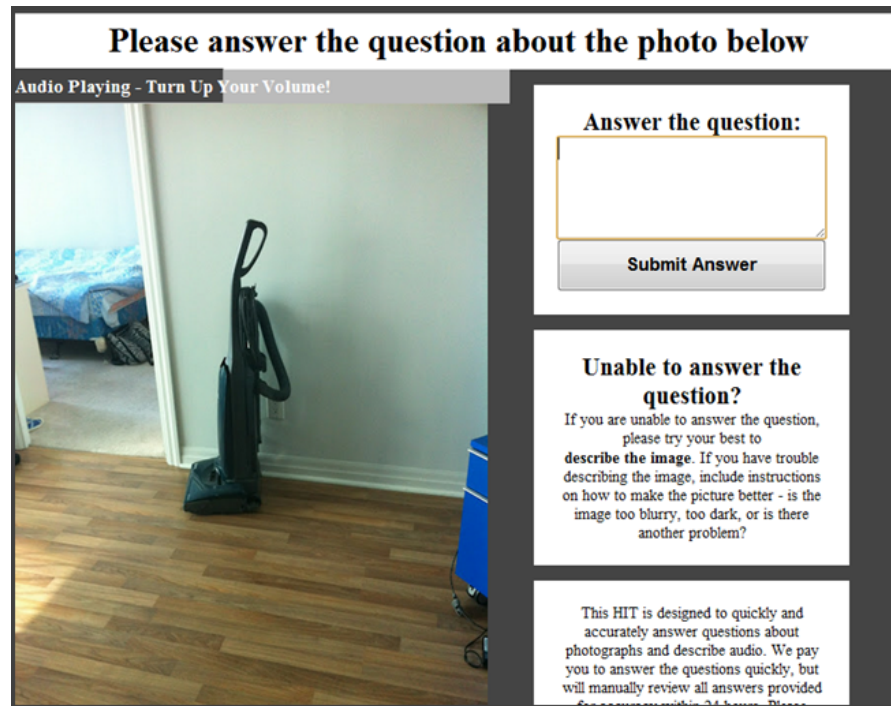
Figure 7.2: The worker interface for people who accept a VizWiz HIT on Mechanical Turk.

users to send their questions to anonymous workers, whom they have no connection with. While the service could maintain a consistent pool of trained staff to answer questions, the price to retain staff members to cover a full 24-hour period would be extremely prohibitive. Additionally, a set pool of always-available workers may be inefficient - in times when few questions were being asked, the workers would be left with nothing to do; in times when many questions were being asked, the pool may not be large enough to handle all the requests at once.

Response times for tasks posted on Mechanical Turk vary, based on the number of workers online at a different time, payment amount offered, and other factors [Chilton et al., 2010]. This can cause difficulties for systems like VizWiz, where the speed of the response is a high priority for the user. To work around this, VizWiz tested two

response models: a pre-recruitment model called *quikTurkit* [Bigham et al., 2010], which rounds up workers preemptively and gives them tasks to complete while waiting to ensure that someone is available when the user finally submits their question; and a consistent pool of workers who were always available to answer.

quikTurkit

quikTurkit, available online at `http://code.google.com/p/quikturkit/`, was built as a way to ensure quick responses from Mechanical Turk workers. The tool has two parts: a method for pre-recruiting workers to have available when questions arrive, and a method of promoting Mechanical Turk HITs to make them more attractive and visible to workers.

The worker queue is tied to the linear procession of the user interface. When a user opens the application to the camera screen, an initial ping is sent to the VizWiz server indicating that the user may be preparing to take a photograph. quikTurkit then posts HITs to Mechanical Turk, filled with previously-asked VizWiz questions. Workers who accept the task will have some time to become familiar with the interface and practice with previous questions while the user takes their picture and records their question. Then, when the question is submitted, the workers are already available to answer, cutting down the latency associated with recruitment.

The speed gains from this strategy are shown in Figure 7.3. By shortening or eliminating the time t_r, quikTurkit cuts down the length of time it takes for each question to be answered, and allows the system designer to avoid maintaining a consistent crowd of workers throughout the day. More details about the implementation and speed gains of quikTurkit can be found in [Bigham et al., 2010].

7.4 Evaluation of VizWiz

VizWiz was evaluated through two pilot studies of the initial application, and by exploring data from its' public release in May 2011.
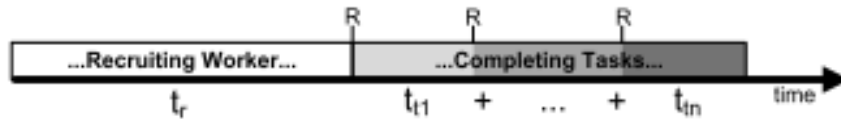
Figure 7.3: The time it takes to recruit workers on Mechanical Turk and have them answer N questions. Recruiting time is represented by t_r, and each of the times to complete tasks are represented by t_{ti}, where $i = 1...N$. quikTurkit and maintaining a pool of workers both provide speed gains, since workers are already recruited, so the total task time is reduced by t_r

7.4.1 Pilot Studies

To evaluate VizWiz, we ran two iterative pilot studies in the field. The first, with no photographic aids and the original version of quikTurkit, ran for 7 days and answered 82 questions from 11 users. Participants were asked to try to use the application once a day or more over the course of the week. Most questions asked for object identification or descriptions of an object or scene, and many (32%) were unanswerable due to blur or darkness in the photographs. Answers took on average 133 seconds, with workers who were pre-recruited as answerers. Participants reported that they enjoyed the system, and thought it would be useful as a widely-available application.

The second field study incorporated blur and darkness detection, and was run with 3 participants for just one day. This field study also used a pool of workers which was maintained throughout the study, with between 4 and 10 workers at any time during the 24 hour period. Maintaining a worker pool was more expensive, but answers came back to the user faster than with prerecruitment - on average, answers took 27 seconds with the worker pool. The cost of maintaining a pool of workers could be prohibitive in the long-term, depending on how frequently users utilize the service.

After the field studies were completed, a full version of VizWiz was built for iOS devices and released on the App Store in May 2011. More recently, an Android version has been developed as well. In the next sections, we discuss the use of VizWiz in the wild, and go on to describe

Figure 7.4: The timeline of a question. On average, it took 61 seconds for users to take their picture, 10 seconds to record their question, and another 7 seconds to send their question. This time constructing questions can be used to pre-recruit workers, and reduce the time required to wait for an answer.

some of the followup projects and spinoffs that have spawned from the initial research.

7.4.2 Questions Asked With VizWiz

As of this writing, over 70,000 questions have been asked with with public version of VizWiz. While some of these questions were expected, and would have been answerable by specialized automated solutions (*e.g.,* identification containers of food in the kitchen, reading the return address of letters received in the mail), others highlight the rich variety of situations in which blind people need visual access to information, and how humans play a valuable role in interpreting and providing that information.

The large number of questions asked with VizWiz has provided us with an unprecedented glimpse into the everyday needs of blind people. Traditional studies, where users are interviewed or asked to keep diaries of their information needs, may be skewed as participants forget certain fleeting question or are subject to recall biases. VizWiz, on the other hand, gives us access to users' in-the-moment needs, and can help us design and build tools to facilitate better access to the information they want.

After VizWiz had been deployed for a year, we analyzed a random sample of 1000 questions asked using the service in order to learn more about how blind people were utilizing the service, and what unmet needs they had. The questions were analyzed on the dimensions of question type, primary photograph subject, perceived question urgency, question subjectivity, and photograph quality. The results in this section are drawn from our paper on this analysis [Brady et al., 2013a].

Question Type

One of the purposes of investigating VizWiz's use was learning what *types* of questions need sighted help to answer. By learning more about the questions users asked, and whether or not VizWiz could successfully answer them now, we could plan future research projects and try to improve both the physical accessibility of a blind person's environment, and the sighted assistance facilitated for them by technological devices.

We devised a categorization of question types through an affinity diagramming process, where questions were silently grouped together by researchers in a first pass categorization; then categories and subcategories were refined and named in a second pass. These categories and subcategories are explained in further depth, with example questions and photographs, in Figure 7.5.

The most common types of questions asked were *identification* questions (41%), where the user asked for the name or type of an object in the photograph. More than half of these questions were simple, asking just "what is this" or "what do you see in the picture", with little contextual details provided about what the users were asking about. Others questions were more detailed (*e.g.,* "what book is this?"). This type of question may have been most popular for many reasons. The VizWiz service was released before other object recognition applications for blind users like TapTapSee[1], so users may not have had other venues to ask these types of questions. Additionally, users may have not wanted to ask their friends or family simple questions - as we will discuss below, blind users may be reluctant to ask their friends and family repeated or trivial questions (at least via social media) for fear of bothering them.

After identification questions, the next most common type was *description* questions (24%), where the user asked for a description of some physical characteristic of an object or environment. Many of these questions asked for color information for objects (24%), and 38% asked about about the users' physical appearance or outfit (their looks, clothing color, or clothing design). These questions demonstrate a desire
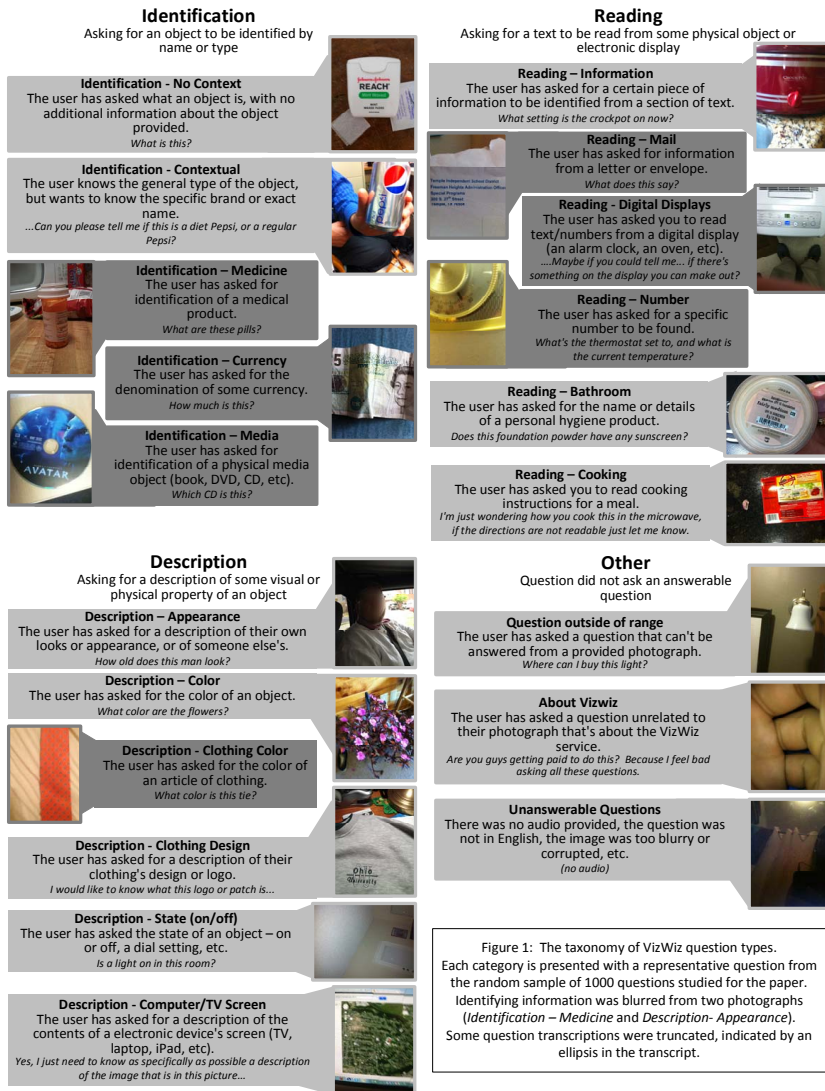
---

[1] http://www.taptapseeapp.com

Identification
Asking for an object to be identified by name or type

Identification - No Context
The user has asked what an object is, with no additional information about the object provided.
What is this?

Identification - Contextual
The user knows the general type of the object, but wants to know the specific brand or exact name.
...Can you please tell me if this is a diet Pepsi, or a regular Pepsi?

Identification – Medicine
The user has asked for identification of a medical product.
What are these pills?

Identification – Currency
The user has asked for the denomination of some currency.
How much is this?

Identification – Media
The user has asked for identification of a physical media object (book, DVD, CD, etc.).
Which CD is this?

Reading
Asking for a text to be read from some physical object or electronic display

Reading – Information
The user has asked for a certain piece of information to be identified from a section of text.
What setting is the crockpot on now?

Reading – Mail
The user has asked for information from a letter or envelope.
What does this say?

Reading - Digital Displays
The user has asked you to read text/numbers from a digital display (an alarm clock, an oven, etc).
....Maybe if you could tell me... if there's something on the display you can make out?

Reading – Number
The user has asked for a specific number to be found.
What's the thermostat set to, and what is the current temperature?

Reading – Bathroom
The user has asked for the name or details of a personal hygiene product.
Does this foundation powder have any sunscreen?

Reading – Cooking
The user has asked you to read cooking instructions for a meal.
I'm just wondering how you cook this in the microwave, if the directions are not readable just let me know.

Description
Asking for a description of some visual or physical property of an object

Description – Appearance
The user has asked for a description of their own looks or appearance, or of someone else's.
How old does this man look?

Description – Color
The user has asked for the color of an object.
What color are the flowers?

Description - Clothing Color
The user has asked for the color of an article of clothing.
What color is this tie?

Description - Clothing Design
The user has asked for a description of their clothing's design or logo.
I would like to know what this logo or patch is...

Description - State (on/off)
The user has asked the state of an object – on or off, a dial setting, etc.
Is a light on in this room?

Description - Computer/TV Screen
The user has asked for a description of the contents of a electronic device's screen (TV, laptop, iPad, etc).
Yes, I just need to know as specifically as possible a description of the image that is in this picture...

Other
Question did not ask an answerable question

Question outside of range
The user has asked a question that can't be answered from a provided photograph.
Where can I buy this light?

About Vizwiz
The user has asked a question unrelated to their photograph that's about the VizWiz service.
Are you guys getting paid to do this? Because I feel bad asking all these questions.

Unanswerable Questions
There was no audio provided, the question was not in English, the image was too blurry or corrupted, etc.
(no audio)

Figure 1: The taxonomy of VizWiz question types. Each category is presented with a representative question from the random sample of 1000 questions studied for the paper. Identifying information was blurred from two photographs (*Identification – Medicine* and *Description- Appearance*). Some question transcriptions were truncated, indicated by an ellipsis in the transcript.

Figure 7.5: Categories and subcategories of questions asked with VizWiz, from [Brady et al., 2013a].

among blind users for assistance with outfit assembly and fashion, a problem area that researchers have begun to study in further depth [Burton et al., 2012, Williams et al., 2013c,b].

Another common type of questions were *reading* questions (17%) where the user asked for textual or numeric information to be read back. While OCR tools exist that can read large chunks of text, many of the questions asked for specific information to be extracted from text or for the sighted worker to use reasoning to report the text the user actually wants (*e.g.,* on a piece of mail, the blind user may ask "who is this from?" - while an OCR tool would read all the text on the envelope, a sighted worker knows just to report the return address from the upper left-hand corner).

Other questions were asked as well. Many of these were uncategorizable, due to poor photography or audio recording, or could not be answered with VizWiz as it currently existed (*e.g.,* a few users asked questions under the impression that VizWiz was a shopping tool, that could identify objects and tell you where to find them in stores).

This initial analysis of VizWiz question types provides a first insight into what assistance blind users need in their daily life. However, the types of questions asked may have been limited by the service itself - for example, while navigation remains an open and widely-examined area of research for blind users [Loomis et al., 1998, Williams et al., 2013a, Jain, 2014], very few VizWiz users asked for navigation advice. This may have been due to the asynchronous nature of the communication (which does not lend itself to navigation, as users must stand still and wait for the next instruction after each question), or the difficulty of answering navigational questions with only photographs and a 15-second recording as input.

Primary Photograph Subjects

While the types of questions asked provide us with inspiration for new tools to build, they do not always give us clear insights into what types of objects or situations are inaccessible for blind users. To examine this in more depth, we also analyzed the primary subject of each photograph taken, to learn if there were trends in the things the questions asked

Sub-category Name	Description	%
Food/Drink	A food product or beverage, either packaged or unpackaged	28%
Computer/TV	The screen of a computer or television, and any accessories (*e.g.,* remote, keyboard, mouse) that go with those devices	8%
Clothing	A object of clothing or accessory, either worn by someone or displayed on a table or hanger	8%
Household	Furniture, appliances, or electronic devices	7%
Entertainment	A toy, craft, or media (*e.g.,* video game, CD, book)	6%
Paper	A letter or piece of paper	6%
Bathroom	Shampoo, conditioner, or other beauty and hygiene products	6%
Miscellaneous	Any object that does not fit into another sub-category	12%

Table 7.1: Primary subject sub-categories and definitions for photographs in the *Object* category, from [Brady et al., 2013a]

about. Due to photographic errors, or uncertainty as to which object was being asked about, primary subjects could only be identified in 88% of questions.

Almost all of the primary subjects identified were *objects* (77%), though some photographs also focused on *settings* or *people or animals*. As shown in Table 7.1, the breakdown of subcategories for the objects asked about in photographs reveals some clear groupings, with food and drink objects being the most commonly asked about.

One bias that may have impacted these results is the marketing of the VizWiz project. On the iTunes Store page where the application could be downloaded, the example question shown was asking about the cooking instructions for a frozen meal. As a result, users may have been unconsciously influenced to ask about cooking or food products

when using the service, or may have used those products as an initial test of its' efficacy.

Perceived Urgency

We also classified the perceived urgency of each question, as a way to determine how quickly questions should be answered. Since users did not have a way to specify how quickly they wanted their question answered, we were not able to base this analysis on their desired urgency level; instead, it was based on researcher's perception of question urgency, and should be considered accordingly. Still, these ratings give us some sense of the importance of speed in our system design - if questions mostly appear to be urgent, then tools like quikTurkit which allow for fast recruitment of workers are essential.

Urgent questions were those in which the rater perceived that the user needed answers quickly, perhaps in order to take a follow-up action like visiting a doctor:

> "Can you tell me, on the top of this baby's head, is there a red rash, or red spots, or do you see anything red or pink or that looks abnormal?" [Brady et al., 2013a]

Non-urgent questions were those in which the rater perceived that the user could wait longer for an answer (*e.g.,* a woman asking what kind of plant was in her garden). 68% of categorizable questions asked were urgent, needing answers within a minute or 1 to 10 minutes.

Subjectivity and Objectivity

Many of the object recognition-type questions are highly objective. For example, if a VizWiz user takes a good-quality photograph of a can of food and asks what it contains, multiple sighted workers can listen to the question and look at the photograph, and they will likely all provide the same answer. However, VizWiz is also useful for subjective questions, where different answers might provide different responses based on their own opinions or biases. Many of the fashion questions referenced above are subjective - one worker might have the same fashion sensibility as the VizWiz user, while another may have completely

opposite tastes. While many objective questions could feasibly be answered by automated technologies in the near future, these types of subjective questions may always require human input for answers to be valuable.

We characterized the objectivity and subjectivity of questions on a 5-point Likert scale. Almost all of the questions that could be categorized were objective (78%). This has implications for the importance of human workers for VizWiz - if many of the questions asked are objective, it is possible that automated tools could be developed to answer them (*e.g.,* a barcode reader could be used to identify commercial products, a color recognizer could describe outfits). However, as the next section describes, the quality of the photographs being taken may not be sufficient for current automated tools.

Photograph Quality

As revealed in the original VizWiz pilot and other related research, many blind users have difficulty with taking photographs. They may take photographs that are too dark or blurry to see at all, or photographs that are not conducive to answering the question that accompanies it. This can cause difficulties for computer-vision tools, which are often trained to work under optimum lighting conditions and with little blur. We examined photographs for the four most common photographic errors we saw (as defined in Brady et al. [2013a]):

Blur: Is the photograph blurry?

Lighting: Is the photograph too dark (*e.g.,* poor lighting in the room) or too bright (*e.g.,* a window or light is directly behind the object)?

Framing: Are parts of the necessary items outside the photograph's edge?

Composition: Is the item obscured by other objects, or by the photographer's finger over the lens?

We gave each picture an initial score of 5, and subtracted 1 point for each photographic error described above (so scores could range from 1 to 5). Only 18% of photographs had 5s, meaning there were no significant issues with the photographs. The average score was 3.41. While many photographs had one or two errors (62%), only 5% of those had unclear subjects in the previous analysis, indicating that answerers were likely still able to answer the questions despite some photographic issues; this may mean that human workers are able to answer questions that computer vision cannot (though a more direct comparison is needed). The photographs taken also provide a valuable data set for training automated tools to support blind photography or adjust computer vision algorithms to accommodate photographs from blind users.

7.5   Conclusions

VizWiz allows access to the physical world by mediating help-seeking through technology. The real-world help-seeking model of asking a friend, family member, or stranger is translated to a remote work context, meaning that workers are available at any time and are able to answer questions quickly.

The privacy implications of VizWiz are concerning, since users do not know if they are revealing information to workers which may be private or sensitive. However, some users intentionally took pictures of sensitive information, like credit cards or mail, because the latency trade-offs of VizWiz were worth the potential risk. This indicates the importance of real-time tools for question answering for blind users, and may motivate future automated tools for these types of tasks.

VizWiz and Social Accessibility (Chapter 6) both make visual information accessible, but VizWiz targets quick, one-off tasks. A consequence of its prerecruiting model for reducing latency is the on-demand availability of groups of workers. This method lead to a new sub-field of human computation called real-time human computation, which has explored a number of interactive tools that employ the power of people [Bernstein et al., 2011, Lasecki et al., 2011, 2012, 2014a]. In the next

chapter (Chapter 8) we explore one of these tools, Scribe [Lasecki et al., 2012], which uses groups of people to convert speech to text for Deaf and Hard of Hearing individuals in real-time (latencies of less than 5 seconds from when a word is spoken to when it appears).

8

Case Study 3: Combining Inputs for Real-Time Captioning with Scribe

Real-time captioning is an accommodation necessary to provide Deaf and hard of hearing people access to spoken content, *e.g.,* lectures, meetings, etc [Cavender et al., 2009]. Computation is still far from being able to handle this task, with even the best offline speech recognizers utilizing deep learning and with vocabularies adapted to the domain only able to achieve word error rates of more than 10% [Hannun et al., 2014]. The error rates rise to 50% or more when used in non-ideal settings, such as exactly those in which Deaf and hard of hearing students would like to use it, *e.g.,* a classroom in which the professor hasn't spent time training the recognizer and isn't wearing a microphone, or an informal meetings or discussions after class.

The only effective way to do real-time captioning is to use stenographers who have trained to use special chorded keyboards to quickly type text as it is spoken. As rare experts, stenographers command wages of $100-300 USD per hour, and generally must be arranged far in advance for blocks of an hour or more. Applying a human-powered approach to this domain is not straightforward because members of the crowd are assumed to not be experts and, therefore, to not have the

training available to type at real-time speaking rates (often over 200 words per minute (wpm), and over 300 wpm in certain bursts).

This chapter presents a case study of Scribe [Lasecki et al., 2012], a real-time captioning system which combines inputs from multiple non-experts to convert speech to text with very low latency (less than 5 seconds). Scribe asks multiple non-expert workers to contribute what they can and deeply integrates their work with computation to allow them to collectively caption at real-time speaking rates. By combining these inputs, the crowd can serve as an expert captioner, without the traditional overhead of training and scheduling.

8.1 Scribe

Scribe is a captioning tool for a variable number of non-expert captioners [Lasecki et al., 2012]. Similar to VizWiz, it requires no special knowledge from the captioners beyond the ability to hear a stream of audio and type. These captioners can be remote workers from crowdsourcing platforms, or even recruited from the audience of the event that is being captioned (*e.g.*, students in a lecture hall can caption their own class, leading to real-time captions for Deaf or hard of hearing classmates, and a transcript of the lecture for themselves afterwards).

A stream of speech is recorded, and is transcribed by multiple workers. Each worker is asked to transcribe a small, manageable portion of the audio, rather than attempting to have each worker transcribe the entire stream, but is able to listen to the rest of the audio in order to gain context from the rest of the speech. The portions captioned by each worker are then "stitched" together computationally using a variation of Multiple Sequence Alignment designed to work in near real-time over word sequences [Naim et al., 2013].

Various improvements have been made to make the captioning process easier for non-experts, like slowing down the sections of audio that each person is transcribing [Lasecki et al., 2013]. This surprisingly not only results in higher accuracy, but also lower latency as workers no longer need to remember what they heard and type it later when the audio is at or below their typing speed.
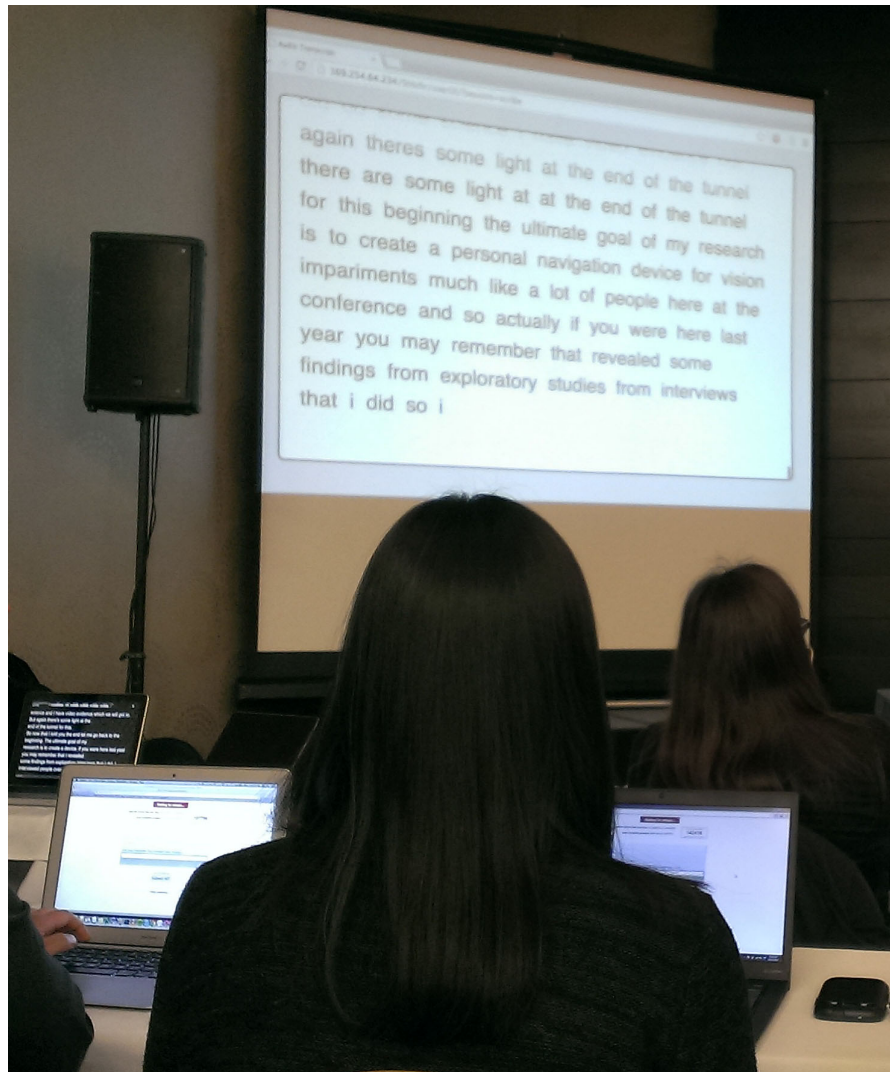
Figure 8.1: A photo of Scribe in use at the ACM ASSETS 2014 conference. In this demonstration of the system, five audience members were asked shortly before an upcoming session to help collectively caption it using Scribe. The photo shows the captioning interface on one of the volunteers' laptops and the displayed captions in the background.

Captioners can all see the stream of captions, allowing them to edit mistakes in the alignment process, or see homophones that occur in the speech stream and repeat them in their own transcriptions. The regularity of words produced is higher and the latency approaches and can even exceed that of expert stenographers [Lasecki et al., 2012]. Scribe's flexibility has allowed for advances in the interfaces used to view the captioning [Lasecki et al., 2014c], and also allowed for different approaches looking at how automatic speech recognition may lower (and, perhaps, eventually replace) the need for human workers.

8.2   Design Dimensions of Scribe

The design dimensions presented in Chapter 5 can be applied to Scribe, for a better understanding of how it fits into the landscape of human-powered access tools.

8.2.1   Users

Scribe is designed for Deaf and hard-of-hearing users, who may need real-time captioning to follow along in classrooms or meetings. However, it can be useful to others as well, as students in those classes will then have a transcription they can use during studying, or colleagues can refer back to the transcript of the meeting to review agenda items or to-dos. The initiative of the system is either the end user or organizations - individuals can start recording speech around them to be transcribed, but in more formal lecture settings the system would likely be initiated by a professor or the university.

When used in formal presentations, there are few broader considerations which need to be made, since speakers understand they are in a public setting and being observed. However, if used in general situations, people may be unaware they are being recorded. There are no features in place to indicate that the system may be recording sensitive information, leaving it to the user or initiator to decide if recording is inappropriate at any time.

8.2.2   Tasks

Volunteers could caption the audio streams for free, while Mechanical Turk workers must be paid (by either the individual or an organization) commensurate to the time they spend on the task. When recruiting from Mechanical Turk, it is likely there will always be enough workers available if the compensation is high enough, ensuring that the system can be up and running quickly and nearly always available. The latency of the system is expected to be near-immediate (within a few seconds), and as a result there are no accuracy guarantees from the system beyond the redundancy inherent in the caption combination algorithm.

The real-time captions can be used by anyone present for the event, or even for video streams or stand-alone live transcripts. However, the transcripts can also be accessed after the fact, allowing for both individual and group reuse. These transcripts could also be integrated into ASR tools, used as a training set to improve the quality of speech recognition in scientific reuse.

8.2.3   Workers

As discussed above, the system can be powered by either organized volunteers or paid crowd workers, but no expertise is expected or required from either group. Volunteers are motivated by intrinsic factors, or are self-motivated by the promise of the full transcription after the fact. Paid workers are motivated by the financial incentives, and are compensated higher if their captions match highly with other workers.

The assistance provenance is obfuscated to the caption users. Due to the combination algorithm, it is unclear which captions come from which worker, or even how many workers are currently captioning with the system.

8.3   Evaluation of Scribe

Comparing the efficacy of crowd-generated captions to captions from trained transcriptionists or a ground truth transcription can be difficult. If two captions differ only in one word, it is easy to determine

that difference; however, if sections of the text are re-arranged but still present, or differ slightly in structure (*e.g.,* "any one" versus "someone"), it may be harder to compare the two.

To evaluate Scribe, the authors relied on three different metrics: *precision*, the number of words in the caption that appear in the audio stream within a ten second range of their placement in the caption; *coverage*, the number of words in the ground truth transcription that appear in the caption, and *latency*, the average speed in which each word takes to be transcribed. It is interesting to note that speech recognition often produces similar recall as a single human worker, but its precision is so low that its output is difficult to use in conjunction with human input.

The precision of the original version of Scribe (80.3% when inputs from multiple workers are combined) approached, but did not reach the precision of an expert captioner (94.7%). However, the coverage of Scribe increased as more workers were added, eventually surpassing the experts' coverage (88.5%) to reach 93.2% coverage with 10 non-experts. The combined workers were able to achieve better latency (2.89 seconds) than the trained captioner (4.38 seconds) as well. Additions of other features, like the audio warping in [Lasecki et al., 2013], lead to improvements in all three of coverage, precision, and latency. Scribe had better precision, coverage, and latency than automatic speech recognition as well.

8.4 Conclusions

Scribe shows that non-expert workers can approximate (and, in the cases of coverage and latency, exceed) experts. By designing workflows which make tasks simple for novice workers, and leveraging that input against each other, crowd-powered systems can complete tasks previously thought to require specialized workers. In the case of Scribe, this transforms real-time captioning from a costly and complicated process to a simple one.

Scribe presents a compelling hybrid of volunteer and paid labor. Volunteers are able to benefit from the system, and may be encour-

aged to participate for intrinsic or self-motivated reasons; however, when enough volunteers are not available, the system can recruit workers from Mechanical Turk at little loss of quality. Volunteers who are familiar with the source material may have greater domain knowledge and be able to identify specialized terms, but the Turk workers can focus fully on the transcription task without having their attention split between transcribing and listening.

While Scribe is similar to VizWiz (Chapter 7) in its nearly-realtime approach to providing access to the physical world, the task type is significantly different. VizWiz questions are one-off questions, which are meant to be answered by a worker and then the task is complete; Scribe requires human input over a sustained period of time. Since Mechanical Turk workers and volunteers can not be guaranteed to stay available for the entire transcription task, the workflow is designed to accommodate workers joining and abandoning the task, reducing the burden on any one Turker and providing a mechanism for dealing with worker turnover.

Scribe demonstrates two general possibilities for human-powered access technology. First, it shows that tools with very low latency constraints that need to be operated continuously can be powered by the crowd. A number of tools have been developed with this approach, both to support accessibility and for general use [Lasecki et al., 2014b, 2011, 2014a, 2015]. Second, it demonstrates the ability of multiple non-experts working in conjunction to perform the task of an expert.

9

Challenges And Opportunities Going Forward

Having examined a number of human-powered access technologies and their dimensions, we can see the power of remote people as a source of information or assistance for people with disabilities. These tools can enable users to access information more independently and complete tasks without another person physically accompanying them.

As more tools are developed and technological advancements are made, new challenges and opportunities for human-powered access will arise. In this chapter, we discuss these concerns, and how we believe they can be leveraged or mitigated.

9.1 Challenges

9.1.1 Costs

One of the primary challenges, even in current crowdsourcing approaches, is the financial cost of maintaining a service. These costs occur in several domains: for the system designers, both in recruiting and retaining workers and in maintaining the service; for the user, in the initial cost of purchasing the device(s) to facilitate assistance, and on-going costs related to use.

System Costs

The financial costs of recruiting workers for services have been briefly discussed above, but cannot be discounted when designing access technologies. Retained worker pools cost money and may not be used at predictable intervals, while individually-recruited workers may need to be compensated at higher rates for small tasks, as they need to learn the task and interface before getting started.

The community of researchers and practitioners who are crowdsourcing has grown greatly since Mechanical Turk was introduced in 2005, and there may be more competition for a limited number of workers that will drive up costs. Additionally, as requesters' understanding of Mechanical Turk and other marketplaces has shifted from that of a place where people do small tasks for fun or for pocket money to that of a labor market, more Turkers and requesters have begun to call for ethical payment rates which are comparable to the national minimum wage.

Other costs arise in maintenance of the system. Due to the ever-changing nature of web APIs and technological devices, developers may be frequently required to update their applications or repair integrations with other products. The services themselves will likely cost money to maintain, with servers to host media or content produced by the users, and developer licenses or fees for releasing tools be charged on an annual basis.

User Costs

The user may need to pay expensive initial costs in order to utilize human-powered access technologies at first. For example, while VizWiz provides answers to visual questions for free, until recently the service was only available on the iPhone, which limited use to people with a phone plan that included Apple devices, or to those who could afford the devices ($200 or more). While the cost of an iPhone is significantly lower than specialized devices (which are often $1000 or more), and the phone can be used for other purposes besides assistance, this cost may still be out of reach for many users. Other access devices may require

expensive technology as well, such as video recording for navigational tasks or audio recording for transcription tasks.

Other costs may be directly related to use of the service. Some services, like TapTapSee, operate on a paid model where users must pay for each question they ask, or subscribe on a monthly basis. Some services may not charge money directly, but instead consume bandwidth and the users' data plan by transmitting large amounts of data (like video files).

While the costs of crowdsourcing may appear cheap, quality verification methods can lead to additional expenses. For users with disabilities, this can lead to a tradeoff of costs vs. high-quality answers. As the case study of Solona exhibited, many of these services can be invaluable to users, and it may be worth having lower-quality or unverified answers if it means the service is able to be maintained for longer on the same costs. New tools, like BeMyEyes, are moving to more cost-effective models by using volunteer workers, which keeps users with disabilities from having to choose between quality and expenses.

9.1.2   Starting a New Service

In the beginning of the service, it may be hard to know what type of worker recruitment will be necessary for the system to fill all users' needs. Tradeoffs between maintaining a large enough worker pool when the system is used by few people, and maintaining enough workers for large numbers of users who may try out the system simultaneously when they hear or read about it, may be difficult to anticipate and influence the user's first impressions.

The method in which the system is distributed to potential users may also impact its uptake and eventual success. Systems can be distributed through organizations that work with people with disabilities, distributed to friends or colleagues who may share with potential users, or released with little fanfare for users to search out themselves. While not publicizing a service and leaving it to be sought out by potential users could seem counter-intuitive, it may allow the system to scale slowly, and prevent the system from being misused.

9.1.3   Sustainability and Longevity

After a service has started, and gained enough user activity to justify continuing it, the developers may encounter difficulties in maintaining the service for a long time and in a sustainable way.

User Dependence on the Service

Concerns with raising enough funds to continue a service indefinitely, or maintaining fast response speeds when usage of a service may fluctuate, can also impact the users of a service. While mobile and access technologies can make users feel more independent [Kane et al., 2009], if those services are suddenly shuttered, it can have a greater negative impact on the user than if the service never existed.

This was the case with Solona, the CAPTCHA-recognizing service which shut down due to insufficient staffing. After it stopped working, former users took to message boards trying to figure out what was going on and voicing their frustration [Solona, a]:

> *typical! the one service that really solved the problems we have with captchas, and it has to get screwed up!*

The users then began exchanging other methods of solving CAPTCHAs with remote assistance, but none seemed as simple as Solona had been. Making reliable services that don't disappear overnight and that users can trust to exist in the long-term, is another major challenge of human-powered access technologies, and may be mitigated by encouraging competition so that users have alternatives if a particular system cannot be sustained.

Maintaining a Worker Pool

If workers are paid to participate in human-powered access services, then having people available to staff the services indefinitely is simple. While the service must be financially supported, there are many people working on human intelligence marketplaces, and there will always be people looking to earn extra money who can serve as workers.

However, for systems that use non-financial incentives, or that require domain experts, it may be more difficult to retain workers throughout the lifetime of a service.

Workers who volunteer for human-powered access services may have strong intrinsic motivations, like altruism or a desire for social status. However, these motivations are not necessarily enough to keep them engaged with the service. They may have jobs or other commitments which leave them with little free time, or their use of the service may naturally wane over time. Another issue with maintaining the worker pool is that as more services are developed, the demands on volunteer time may increase. If the pool of volunteer labor is not enough to sustain multiple tools, all will have trouble recruiting people to work on them and the market may become oversaturated. One promising approach seems to be expanding the efforts of people interested in a cause by allowing them to easily engage their networks in supporting services they care about (Brady et al. [2015]).

Even with paid workers, where a large pool of people can be reasonably assumed to be available at any time, usage of the service may experience natural or periodic fluctuations, which make it difficult to anticipate how to continue staffing them. While systems that have very few users are likely not financially sustainable, systems with many users may cause other stresses, and may be expensive or experience service delays or outages as a result. Private crowds, such as those run by large corporations or sponsored by government agencies as relay services are, may be able to address many of these issues by subsidizing the costs of human-powered access tools.

9.2 Opportunities

While many unsolved challenges exist in human-powered access technologies, there is also a wealth of opportunities for researchers and developers to exploit.

The increased availability of workers, at all times of day and in large groups, facilitates new types of interactions. Systems like Scribe would not have been possible with one-to-one assistance, and other tools can

follow this model to perform new tasks that were thought of as too hard for humans.

Another exciting opportunity of these systems is enabling diverse and non-traditional workers to participate in the labor force. As seen in the examination of disabled crowd workers by Zyskowski et al. [2015], workers who are often excluded from the labor force are able to do crowdwork due to its flexibility and online nature. This early work presents an intriguing opportunity - could people with disabilities form a self-supporting crowdsourcing system, where people use their unique abilities to help others who cannot access information? While blind workers cannot answer visual questions, they can easily participate in transcription tasks; similarly, Deaf workers could easily do visual identifications.

Researchers can also take advantage of the previously-unstudied accessibility problems that are being identified. Systems that collect user data, like VizWiz, can be used to train object recognition, or to learn about specialized tools that should be developed for certain scenarios (*e.g.,* fashion or cooking). Additionally, researchers can examine message boards or the social web as a way to identify accessibility problems that, in the past, may not have been obvious to people without disabilities. By learning more about existing needs, researchers can avoid wasting time on tools that are unnecessary, and focus on areas where a need truly exists.

10

Conclusions

In this article, we have outlined the dimensions of human-powered access technologies. While people with disabilities often make use of assistance from companions and others around them when they have an access problem, creating scalable technology to replicate and extend this type of interaction with remote assistants offers a number of advantages. Doing so *effectively*, however, involves making a number of decisions and is more complicated than it appears. Not only are more people potentially involved, each with their own expectations and needs, but their contributions must be effectively coordinated, managed, and rewarded.

We have attempted to not only identify these many challenges, but also to highlight opportunities that technologies create for improved accessibility. Systems like Scribe, where large crowds of novices are gathered to do transcriptions, enables real-time access to speech which would not be possible with a traditional companion who is co-located with the user. Bookshare lets individual volunteers scan books into the system, but compiles them into a database that any registered user can access, increasing the scope of books available to any person. These systems are vastly different in how they operate, how they incorporate

human assistance, their incentives for contribution, and the extent to which they use computation; but they share the idea of people coming together, often remotely, to work to improve access.

The types of workers and their motivations for participating can have a significant effect on the speed, correctness, and depth of responses received. While financial incentives are an easy way to seed systems with workers, many projects are operated completely by volunteers, or a mix of paid workers and automation. Other design dimensions mentioned in Chapter 5 can be consulted when thinking about how to design a new tool and working through the constraints that need to be satisfied in a new domain.

By studying examples of human-powered access technologies, and categorizing the design dimensions represented in each, we aimed to provide a deeper understanding of how technology can still be used to increase access in situations where automated tools cannot be used. While our hope is that many of the services powered by human input may one day be automated, and that the data collected by human-powered access technology may even hasten that transition, examples like Social Accessibility, VizWiz, and Scribe show how far we still are from completely automated solutions. Human assistance is certain to continue to play a large role in effective access technologies, not only in the near future but for decades to come, just as human assistance has been key for accessibility for centuries.

Acknowledgements

References

Judd Antin and Aaron Shaw. Social desirability bias and self-reports of motivation: a study of amazon mechanical turk in the us and india. In *Proceedings of the 30th Annual ACM Conference on Human Factors in Computing Systems*, CHI '12, pages 2925–2934. ACM, 2012.

Shiri Azenkot, Sanjana Prasain, Alan Borning, Emily Fortuna, Richard E Ladner, and Jacob O Wobbrock. Enhancing independence and safety for blind and deaf-blind public transit riders. In *Proceedings of the 29th Annual ACM Conference on Human Factors in Computing Systems*, CHI '11, pages 3247–3256. ACM, 2011.

Colin Barnes and Geof Mercer. *Independent futures: Creating user-led disability services in a disabling society*. Policy Press, 2006.

Michael S. Bernstein. *Crowd-Powered Systems*. PhD thesis, MIT, 2012.

Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. Soylent: a word processor with a crowd inside. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 313–322. ACM, 2010.

Michael S. Bernstein, Joel Brandt, Robert C. Miller, and David R. Karger. Crowds in two seconds: enabling realtime crowd-powered interfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 33–42, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0716-1. . URL http://doi.acm.org/10.1145/2047196.2047201.

Jeffrey P. Bigham. Increasing web accessibility by automatically judging alternative text quality. In *Proceedings of the 12th International Conference on Intelligent User Interfaces*, IUI '07, pages 349–352. ACM, 2007.

Jeffrey P. Bigham. Making the web easier to see with opportunistic accessibility improvement. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 117–122, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-3069-5. . URL `http://doi.acm.org/10.1145/2642918.2647357`.

Jeffrey P. Bigham and Anna C. Cavender. Evaluating existing audio captchas and an interface optimized for non-visual use. In *Proceedings of the 27th Annual ACM Conference on Human Factors in Computing Systems*, CHI '09, pages 1829–1838. ACM, 2009.

Jeffrey P. Bigham and Richard E. Ladner. Accessmonkey: a collaborative scripting framework for web users and developers. In *Proceedings of the 2007 International Cross-Disciplinary Conference on Web Accessibility*, W4A '07, pages 25–34. ACM, 2007.

Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23th Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 333–342. ACM, 2010.

Jeffrey P. Bigham, Richard E Ladner, and Yevgen Borodin. The design of human-powered access technology. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '11, pages 3–10. ACM, 2011.

Jeffrey P. Bigham, Michael S. Bernstein, and Eytan Adar. Human computer interaction and collective intelligence. 2014.

John Brabyn, William Crandall, and William Gerrey. Remote reading systems for the blind: A potential application of virtual presence. In *Proceedings of the 14th Annual IEEE International Conference of Engineering in Medicine and Biology Society*, volume 4, pages 1538–1539. IEEE, 1992.

Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. Visual challenges in the everyday lives of blind people. In *Proceedings of the 31rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '13, pages 2117–2126. ACM, 2013a.

Erin Brady, Meredith Ringel Morris, and Jeffrey P. Bigham. Gauging receptiveness to social microvolunteering. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, New York, NY, USA, 2015. ACM. URL `http://research.microsoft.com/pubs/238319/socialmicrovolunteering_chi2015.pdf`.

Erin L. Brady, Yu Zhong, Meredith Ringel Morris, and Jeffrey P. Bigham. Investigating the appropriateness of social network question asking as a resource for blind users. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '13, pages 1225–1236, New York, NY, USA, 2013b. ACM. ISBN 978-1-4503-1331-5. . URL `http://doi.acm.org/10.1145/2441776.2441915`.

Michele A Burton, Erin Brady, Robin Brewer, Callie Neylan, Jeffrey P. Bigham, and Amy Hurst. Crowdsourcing subjective fashion advice using vizwiz: challenges and opportunities. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 135–142. ACM, 2012.

Stefan Carmien, Rogerio DePaula, Andrew Gorman, and Anja Kintsch. Increasing workplace independence for people with cognitive disabilities by leveraging distributed cognition among caregivers and clients. In *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*, GROUP '03, pages 95–104. ACM, 2003.

Anna Cavender, Shari Trewin, and Vicki Hanson. General writing guidelines for technology and people with disabilities. *ACM SIGACCESS Accessibility and Computing Newsletter*, (92):17–22, 2008.

Anna C. Cavender, Jeffrey P. Bigham, and Richard E. Ladner. Classinfocus: enabling improved visual attention strategies for deaf and hard of hearing students. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '09, pages 67–74, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-558-1. . URL `http://doi.acm.org/10.1145/1639642.1639656`.

Anna C. Cavender, Daniel S. Otero, Jeffrey P. Bigham, and Richard E Ladner. Asl-stem forum: Enabling sign language to grow through online collaboration. In *Proceedings of the 28th Annual ACM Conference on Human Factors in Computing Systems*, CHI '10, pages 2075–2078. ACM, 2010.

Dana Chandler and Adam Kapelner. Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, 90:123–133, 2013.

James I. Charlton. *Nothing about us without us: Disability oppression and empowerment.* University of California Press, 1998.

Charles L. Chen and T.V. Raman. Axsjax: A talking translation bot using google im: Bringing web-2.0 applications to life. In *Proceedings of the 2008 International Cross-Disciplinary Conference on Web Accessibility*, W4A '08, pages 54–56. ACM, 2008.

Lydia B. Chilton, John J. Horton, Robert C. Miller, and Shiri Azenkot. Task search in a human computation market. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 1–9, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0222-7. . URL `http://doi.acm.org/10.1145/1837885.1837889`.

Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. Shepherding the crowd yields better work. In *Proceedings of the 15th ACM Conference on Computer Supported Cooperative Work*, CSCW '12, pages 1013–1022. ACM, 2012.

Julie S Downs, Mandy B Holbrook, Steve Sheng, and Lorrie Faith Cranor. Are your participants gaming the system?: screening mechanical turk workers. In *Proceedings of the 28th Annual ACM Conference on Human Factors in Computing Systems*, CHI '10, pages 2399–2402. ACM, 2010.

Jan C Galvin and Marcia J Scherer. *Evaluating, Selecting, and Using Appropriate Assistive Technology.* ERIC, 1996.

Stephanie Hackett, Bambang Parmanto, and Xiaoming Zeng. Accessibility of internet websites through time. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility*, number 77-78, pages 32–39. ACM, 2004.

Awni Y. Hannun, Carl Case, Jared Casper, Bryan C. Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, and Andrew Y. Ng. Deep speech: Scaling up end-to-end speech recognition. *CoRR*, abs/1412.5567, 2014. URL `http://arxiv.org/abs/1412.5567`.

Jonathan Holman, Jonathan Lazar, Jinjuan Heidi Feng, and John D'Arcy. Developing usable captchas for blind users. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '07, pages 245–246. ACM, 2007.

Hwajung Hong, Jennifer G. Kim, Gregory D. Abowd, and Rosa I. Arriaga. Designing a social network to support the independence of young adults with autism. In *Proceedings of the 15th ACM Conference on Computer Supported Cooperative Work*, CSCW '12, pages 627–636. ACM, 2012.

John J. Horton. *Online labor market.* Springer, 2010.

World Health Organization ICF. Towards a common language for functioning, disability and health, 2002.

Panagiotis G. Ipeirotis. Mechanical turk: The demographics, 2008. URL `http://www.behind-the-enemy-lines.com/2008/03/mechanical-turk-demographics.html`.

Panagiotis G. Ipeirotis. The new demographics of mechanical turk, 2010. URL `http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html`.

Panagiotis G Ipeirotis, Foster Provost, and Jing Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 64–67. ACM, 2010.

Lilly C Irani and M Silberman. Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the 31st Annual ACM Conference on Human Factors in Computing Systems*, CHI '13, pages 611–620. ACM, 2013.

Dhruv Jain. Path-guided indoor navigation for the visually impaired using minimal building retrofitting. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '14. ACM, 2014.

Chandrika Jayant, Matt Renzelmann, Dana Wen, Satria Krisnandi, Richard Ladner, and Dan Comden. Automated tactile graphics translation: in the field. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '05, pages 75–82. ACM, 2007.

Chandrika Jayant, Hanjie Ji, Samuel White, and Jeffrey P. Bigham. Supporting blind photography. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '11, pages 203–210. ACM, 2011.

Shaun K. Kane, Chandrika Jayant, Jacob O. Wobbrock, and Richard E. Ladner. Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '09, pages 115–122. ACM, 2009.

Shashank Khanna, Aishwarya Ratan, James Davis, and William Thies. Evaluating and improving the usability of mechanical turk for low-income workers in india. In *Proceedings of the 1st ACM Symposium on Computing for Development*, DEV '10, page 12. ACM, 2010.

Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. The future of crowd work. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '13, pages 1301–1318, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1331-5. . URL `http://doi.acm.org/10.1145/2441776.2441923`.

Richard L. Kline and Ephraim P. Glinert. Improving gui accessibility for people with low vision. In *Proceedings of the 13th Annual ACM Conference on Human Factors in Computing Systems*, CHI '95, pages 114–121, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co. ISBN 0-201-84705-1. . URL `http://dx.doi.org/10.1145/223904.223919`.

Raymond Kurzweil. The Ray Kurzweil Reader, 2003. URL `http://www.kurzweilai.net/pdf/RayKurzweilReader.pdf`.

Richard E. Ladner, Melody Y. Ivory, Rajesh Rao, Sheryl Burgstahler, Dan Comden, Sangyun Hahn, Matthew Renzelmann, Satria Krisnandi, Mahalakshmi Ramasamy, Beverly Slabosky, Andrew Martin, Amelia Lacenski, Stuart Olsen, and Dmitri Groce. Automating tactile graphics translation. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '05, pages 150–157. ACM, 2005.

Walter S. Lasecki, Kyle I. Murray, Samuel White, Robert C Miller, and Jeffrey P. Bigham. Real-time crowd control of existing interfaces. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, UIST '11, pages 23–32. ACM, 2011.

Walter S. Lasecki, Christopher D. Miller, Adam Sadilek, Andrew Abumoussa, Donato Borrello, Raja Kushalnagar, and Jeffrey P. Bigham. Real-time captioning by groups of non-experts. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 23–34. ACM, 2012.

Walter S. Lasecki, Christopher D. Miller, and Jeffrey P. Bigham. Warping time for more effective real-time crowdsourcing. In *Proceedings of the 31st Annual ACM Conference on Human Factors in Computing Systems*, CHI 2013, pages 2033–2036. ACM, 2013.

Walter S. Lasecki, Mitchell Gordon, Danai Koutra, Malte F. Jung, Steven P. Dow, and Jeffrey P. Bigham. Glance: Rapidly coding behavioral video with the crowd. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 551–562, New York, NY, USA, 2014a. ACM. ISBN 978-1-4503-3069-5. . URL `http://doi.acm.org/10.1145/2642918.2647367`.

Walter S. Lasecki, Chris Homan, and Jeffrey P. Bigham. Architecting real-time crowd-powered systems. *Human Computation Journal*, September 2014b.

Walter S. Lasecki, Raja Kushalnagar, and Jeffrey P. Bigham. Helping students keep up with real-time captions by pausing and highlighting. In *Proceedings of the 11th Web for All Conference*, W4A '14, pages 39:1–39:8, New York, NY, USA, 2014c. ACM. ISBN 978-1-4503-2651-3. . URL `http://doi.acm.org/10.1145/2596695.2596701`.

Walter S. Lasecki, Juho Kim, Nicholas Rafter, Onkur Sen, Jeffrey P. Bigham, and Michael S. Bernstein. Apparition: crowdsourced user interfaces that come to life as you sketch them. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, New York, NY, USA, 2015. ACM. URL `http://www.cs.cmu.edu/~jbigham/pubs/pdfs/2015/apparition.pdf`.

Jonathan Lazar, Alfreda Dudley-Sponaugle, and Kisha-Dawn Greenidge. Improving web accessibility: a study of webmaster perceptions. *Computers in Human Behavior*, 20(2):269–288, 2004.

Jonathan Lazar, Aaron Allen, Jason Kleinman, and Chris Malarkey. What frustrates screen reader users on the web: A study of 100 blind users. *International Journal of Human-Computer Interaction*, 22(3):247–269, 2007.

Kimberly Ling, Gerard Beenen, Pamela Ludford, Xiaoqing Wang, Klarissa Chang, Xin Li, Dan Cosley, Dan Frankowski, Loren Terveen, Al Mamunur Rashid, et al. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication*, 10(4):00–00, 2005.

Greg Little, Lydia B Chilton, Max Goldman, and Robert C Miller. Turkit: human computation algorithms on mechanical turk. In *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, UIST '10, pages 57–66. ACM, 2010.

Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. Navigation system for the blind: Auditory display modes and guidance. *Presence: Teleoperators and Virtual Environments*, 7(2):193–203, 1998.

Jennifer Mankoff, Gillian R. Hayes, and Devva Kasnitz. Disability studies as a source of critical inquiry for the field of assistive technology. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '10, pages 3–10. ACM, 2010.

Marc Marschark, Greg Leigh, Patricia Sapere, Denis Burnham, Carol Convertino, Michael Stinson, Harry Knoors, Mathijs P.J. Vervloed, and William Noble. Benefits of sign language interpreting and text alternatives for deaf students' classroom learning. *Journal of Deaf Studies and Deaf Education*, 11(4):421–437, 2006.

David Martin, Benjamin V Hanrahan, Jacki O'Neill, and Neha Gupta. Being a turker. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '14, pages 224–235. ACM, 2014.

João Martins, José Carilho, Oliver Schnell, Carlos Duarte, Francisco M Couto, Luís Carriço, and Tiago Guerreiro. Friendsourcing the unmet needs of people with dementia. In *Proceedings of the 11th Web for All Conference*, W4A '14, page 35. ACM, 2014.

Tara Matthews, Scott Carter, Carol Pai, Janette Fong, and Jennifer Mankoff. Scribe4me: Evaluating a mobile sound transcription tool for the deaf. In *Proceedings of the 8th International Conference of Ubiquitous Computing*, UbiComp '06, pages 159–176. Springer, 2006.

Shigeki Miyoshi, Hayato Kuroki, Sumihiro Kawano, Mayumi Shirasawa, Yasushi Ishihara, and Masayuki Kobayashi. *Support technique for real-time captionist to use speech recognition software.* Springer, 2008.

Iftekhar Naim, Daniel Gildea, Walter S. Lasecki, and Jeffrey P. Bigham. Text alignment for real-time crowd captioning. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Conference*, NAACL '13, pages 201–210, 2013.

Alan F. Newell, Peter Gregor, Maggie Morgan, Graham Pullin, and Catriona Macaulay. User-sensitive inclusive design. *Universal Access in the Information Society*, 10(3):235–243, 2011.

Oded Nov, Ofer Arazy, and David Anderson. Technology-mediated citizen science participation: A motivational model. In *The 5th International AAAI Conference on Web and Social Media*, ICWSM '11, 2011.

Abiodun Olalere and Jonathan Lazar. Accessibility of us federal government home pages: Section 508 compliance and site accessibility statements. *Government Information Quarterly*, 28(3):303–309, 2011.

Ellie Pavlick, Matt Post, Ann Irvine, Dmitry Kachaev, and Chris Callison-Burch. The language demographics of amazon mechanical turk. *Transactions of the Association for Computational Linguistics*, 2:79–92, 2014.

Mark Pilgrim. *Greasemonkey Hacks: Tips & Tools for Remixing the Web with Firefox.* " O'Reilly Media, Inc.", 2005.

Alexander J. Quinn and Benjamin B. Bederson. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the 29th Annual ACM Conference on Human Factors in Computing Systems*, CHI '11, pages 1403–1412. ACM, 2011.

M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review*, 9(1):010103, 2010.

Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. In *The 5th International AAAI Conference on Web and Social Media*, ICWSM '11, 2011.

Joel Ross, Lilly Irani, M Silberman, Andrew Zaldivar, and Bill Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th Annual ACM Conference on Human Factors in Computing Systems*, CHI '10, pages 2863–2872. ACM, 2010.

Niloufar Salehi, Lilly C. Irani, Michael S. Bernstein, Ali Alkhatib, Eva Ogbe, Kristy Milland, and Clickhappier. We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1621–1630, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3145-6. . URL `http://doi.acm.org/10.1145/2702123.2702508`.

Daisuke Sato, Hironobu Takagi, Masatomo Kobayashi, Shinya Kawanaka, and Chieko Asakawa. Exploratory analysis of collaborative web accessibility improvement. *ACM Transactions on Accessible Computing (TACCESS)*, 3 (2):5, 2010.

Marcia Joslyn Scherer. *Living in the state of stuck: How technology impacts the lives of persons with disabilities*. Brookline Books, 1996.

Tom Shakespeare and Nicholas Watson. The social model of disability: an outdated ideology? *Exploring theories and expanding methodologies: Where we are and where we need to go*, (2):9–28, 2001.

Ben Shneiderman. The limits of speech recognition. *Communications of the ACM*, 43(9):63–65, 2000.

Solona. What's happened to solona?, a. URL `https://www.freelists.org/post/access-uk/whats-happened-to-solona`.

Solona. Solona blog, b. URL `http://web.archive.org/web/20091220221455/http://www.solona.net/wordpress/2009/08/important-update-for-solona-users-please-read`. Accessed via the Internet Archive WaybackMachine.

Michael S. Stinson, Sandy Eisenberg, Christy Horn, Judy Larson, Harry Levitt, and Ross Stuckless. *Real-time speech-to-text services*. Rochester Institute of Technology, National Technical Institute for the Deaf, Northeast Technical Assistance Center, 1999.

Molly Follette Story. Maximizing usability: the principles of universal design. *Assistive Technology*, 10(1):4–12, 1998.

Hironobu Takagi, Shinya Kawanaka, Masatomo Kobayashi, Takashi Itoh, and Chieko Asakawa. Social accessibility: achieving accessibility through collaborative metadata authoring. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '08, pages 193–200. ACM, 2008.

Hironobu Takagi, Shinya Kawanaka, Masatomo Kobayashi, Daisuke Sato, and Chieko Asakawa. Collaborative web accessibility improvement: challenges and possibilities. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '09, pages 195–202. ACM, 2009.

Gregg Vanderheiden. Fundamental principles and priority setting for universal usability. In *Proceedings on the 2000 Conference on Universal Usability*, CUU '00, pages 32–37. ACM, 2000.

Marynel Vázquez and Aaron Steinfeld. Helping visually impaired users properly aim a camera. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, pages 95–102. ACM, 2012.

Fernanda B. Viegas, Martin Wattenberg, Jesse Kriss, and Frank Van Ham. Talk before you type: coordination in wikipedia. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, HICSS '07, pages 78–78. IEEE, 2007.

Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the 22nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '04, pages 319–326. ACM, 2004.

Luis Von Ahn, Manuel Blum, and John Langford. Telling humans and computers apart automatically. *Communications of the ACM*, 47(2):56–60, 2004.

Luis Von Ahn, Shiry Ginosar, Mihir Kedia, Ruoran Liu, and Manuel Blum. Improving accessibility of the web with a computer game. In *Proceedings of the 24th Annual ACM Conference on Human Factors in Computing Systems*, CHI '06, pages 79–82. ACM, 2006.

William Yang Wang, Dan Bohus, Ece Kamar, and Eric Horvitz. Crowdsourcing the acquisition of natural language corpora: Methods and observations. In *Proceedings of the Spoken Language Technology Workshop*, SLT '12, pages 73–78, 2012.

Michele A Williams, Amy Hurst, and Shaun K Kane. Pray before you step out: describing personal and situational blind navigation behaviors. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, page 28. ACM, 2013a.

Michele A Williams, Callie Neylan, and Amy Hurst. Preliminary investigation of the limitations fashion presents to those with vision impairments. *Fashion Practice: The Journal of Design, Creative Process & the Fashion Industry*, 5(1):81–106, 2013b.

Michele A Williams, Kathryn Ringland, and Amy Hurst. Designing an accessible clothing tag system for people with vision impairments. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, page 46. ACM, 2013c.

Jacob O. Wobbrock, Shaun K. Kane, Krzysztof Z. Gajos, Susumu Harada, and Jon Froehlich. Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)*, 3(3):9, 2011.

Haofeng Zhou, Denys Baskov, and Matthew Lease. Crowdsourcing transcription beyond mechanical turk. In *The 1st AAAI Conference on Human Computation and Crowdsourcing*, HCOMP '13, 2013.

Kathryn Zyskowski, Meredith Ringel Morris, Jeffrey P. Bigham, Mary L Gray, and Shaun Kane. Accessible crowdwork? understanding the value in and challenge of microtask employment for people with disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '15, 2015.