Visual Challenges in the Everyday Lives of Blind People

Erin Brady[1], **Meredith Ringel Morris**[2], **Yu Zhong**[1], **Samuel White**[1], **and Jeffrey P. Bigham**[1]
University of Rochester, Computer Science[1]
Rochester, NY 14627 USA
{brady, zyu, jbigham}@cs.rochester.edu
samuel.white@rochester.edu
Microsoft Research[2]
Redmond, WA 98052 USA
merrie@microsoft.com

ABSTRACT
The challenges faced by blind people in their everyday lives are not well understood. In this paper, we report on the findings of a large-scale study of the visual questions that blind people would like to have answered. As part of this year-long study, 5,329 blind users asked 40,748 questions about photographs that they took from their iPhones using an application called VizWiz Social. We present a taxonomy of the types of questions asked, report on a number of features of the questions and accompanying photographs, and discuss how individuals changed how they used VizWiz Social over time. These results improve our understanding of the problems blind people face, and may help motivate new projects more accurately targeted to help blind people live more independently in their everyday lives.

Author Keywords
Blind Users; Q&A; Accessibility; Crowdsourcing; Mobile

ACM Classification Keywords
H.5.m Information Interfaces and Presentation: Misc.

General Terms
Human Factors; Experimentation.

INTRODUCTION
Blind people confront a number of visual challenges every day – from reading the label on a frozen dinner to figuring out if they're at the right bus stop. While many tools have been introduced to help address these problems using computer vision and other sensors (talking OCR, GPS, radar canes, etc.) [18, 16, 6, 5], their capabilities are dictated as much by the state-of-the-art in technology as they are by real human problems. A deeper understanding of the questions that blind people would like to ask in their day-to-day lives may help to direct innovation to solve them.

In this paper, we present the results of a year-long deployment of *VizWiz Social* that provides a new look into the diversity of questions that blind people want answered about their visual environment. VizWiz Social is an iPhone application that lets a blind person take a picture, speak a question they'd like to know about the picture, and then get an answer back within a minute or so from "the crowd" [5]. VizWiz Social has been released "in the wild" since May 2011, and blind users have asked over 40,000 questions since then. Today's technology is targeted at answering some of them, e.g. "What color is this shirt?" and "What does this letter say?" But, others it cannot, for instance, "How many lines are on this pregnancy test?", "What does the sky look like right now?", and "Is my girlfriend hot?"

To help make sense of this diversity, we developed a taxonomy of the questions asked. By outlining the types of questions asked frequently, we hope to improve understanding of the challenges blind people face and help to motivate research into new technology to answer those questions automatically, which would be cheaper and faster. VizWiz Social also provides a rare look into the adoption of an assistive technology over the long term, and how a human-powered access technology [6] affects the user. For instance, do blind people become better photographers as they use VizWiz Social?

VizWiz Social provides insight into a specific but important subset of challenges faced by blind users, i.e., those that can be represented with a still photograph and brief audio description and that can be answered quickly but asynchronously. Other types of challenges, such as those where a user needs help in a situation requiring conveying and/or receiving continuous information, are beyond the bounds of the current study. The pattern – taking a picture and receiving information about it – is also present in much of the automatic technology in use today, and so is a familiar paradigm.

After discussing related work and describing the VizWiz Social application and the data set that we collected through its deployment, we present our analysis of 1,000 questions asked by VizWiz Social users. This unique data set provides insight into the types of challenges faced by blind users and the types of technological solutions that might best address them. We discuss common question types, the nature and urgency of the information being sought, and photographic subjects and quality. We also analyze the behavior of 100 users, to understand what novice and expert interactions with VizWiz Social can teach us about deploying a technology for this audience. We conclude by discussing the implications of our findings on technology design for blind users.

RELATED WORK
Work related to VizWiz Social includes (i) existing mobile access technology used by blind people, (ii) work exploring

CHI 2013, April 27–May 2, 2013, Paris, France.

mobile and social question and answer (Q&A) behaviors, and (iii) systems that use "on-demand" crowd workers.

Mobile Access Technology

For most of the past few decades, mobile access technology for blind people came in the form of specialized hardware devices. Products like talking barcode readers, color identifiers, and talking Optical Character Recognition took the form of expensive, dedicated hardware costing hundreds of dollars [18]. Such devices were limited by the capabilities of automatic technology and had limited uptake.

In the past few years, many standard mobile devices have started to include screen reading software that allows blind people to use them. For instance, Google's Android platform and the Apple iPhone (starting with the 3GS) now include free screen readers [29, 9]. Touchscreen devices like the iPhone were once assumed to be inaccessible to blind users, but well-designed, multitouch interfaces leverage the spatial layout of the screen and can even be preferred by blind people [17]. The iPhone has proven particularly popular among blind users, which is why we developed VizWiz Social for it.

With the availability of an accessible platform, a number of applications were developed for blind people, including GPS navigation applications, OCR readers, and color recognizers. One of the most popular is LookTel[1], which simply identifies U.S. currency denominations (but does so very reliably). Other popular applications help blind people take pictures and identify objects [16].

The capabilities of automatic systems are limited, however, and so several applications have adopted the ideas of "human-powered access technology" [6]. The oMoby application is an accessible application that does object recognition using both computer vision and human computation[2]. VizWiz, from which VizWiz Social derives, explored a more open ended model in which blind people can ask any question about a photo and receive answers back quickly from crowd workers on Mechanical Turk [5].

Mobile Q&A

Mobile devices such as cellular phones enable their users to access people and information from afar. The information people seek in mobile situations has different requirements, usually based on the context of the question-asker, and many services have been developed in order to try to fill these information needs. Because VizWiz Social is an iPhone application, understanding the use of mobile phones for information seeking by general audiences provides important context for interpreting our own findings.

Mobile Information Seeking

The high market penetration of cellular phones has led to them becoming a natural resource for mobile information seeking. 83% of American adults report owning at least one cellular phone, and 51% of those adults have used their cellular phone to access information that they needed right away [27].

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

[2]https://www.iqengines.com/omoby/

As a result of being outside of the home, cellular phone users' *information needs* may differ from regular Q&A needs. A 2009 diary study where users recorded all their mobile information needs indicated that 30% were location-based [11]. Another diary study conducted in 2008 indicated that when a user is outside of their home, the number of information needs that are *contextual* information needs - based on the user's current activity, their location, the time, or a conversation they were involved in - was 72% [28]. However, as smartphones have become more common, "mobile" information seeking has also been performed while the user isn't mobile - a 2011 diary study found that over 70% of "mobile" information seeking was actually performed in familiar, stationary contexts (such as home or the office) [10]. However, this study still indicated that context was a large factor in mobile queries, despite the possibility of the user being in a familiar space.

Mobile Q&A Applications

Non-visual mobile applications were developed in order to answer questions for users while on the go. Google SMS was an automated text-message based tool that allowed users to request information about a limited number of topics (eg. sports scores, weather, word translations). The service could also provide a small amount of information on any topic by sending the user "web snippets" - the first 400-500 characters worth of the first Google search result for that term [13].

Mobile human-backed applications have also become available to answer more complex questions. ChaCha [8] (which bills itself as a "human search engine"), KGB [19], and Naver Mobile Q&A [20] connect users via SMS to human volunteers, who can look up answers to questions for them on the web and send the answers back via SMS. Questions sent to these services are not limited to specific topics and may receive more complex answers than those from Google SMS. These services focus on text-based questions, while VizWiz Social focuses on anwering photographic questions about items in a user's immediate environment.

Some automated mobile applications allow users to get answers to questions about photographs they have taken. Barcode scanners, like RedLaser [25], can identify barcodes from user images and locate relevant product information. In addition to identifying bar codes, the Google Goggles application [12] let users snap photographs of landmarks or works of art and learn about their history, or photograph text and have it read or translated. However, these services relied on computer vision techniques to find and identify objects of interest in the photographs (rather than human workers, as used by our system). As a result, the images needed to be relatively clear and well-centered, and no feedback on how to improve the image was provided to the users if their photographs were not recognizable. This method of interaction may be less suitable for blind users, who cannot view and correct the photographs they are taking.

Social and Crowdsourced Q&A

Our VizWiz Social system allows blind users to send their questions to human sources for answering, either known users (via social networks or email) or anonymous users (via

crowdsourcing services). The use of social and crowd sources for information seeking has been explored, albeit largely in the context of sighted users. For example, Morris et al. [22] found that more than half of users had posted a question to their online social network, and that such questions were largely subjective in nature, seeking opinions and recommendations. Paul et al. [24] studied public Twitter questions and found that rhetorical questions (i.e., those not expecting an answer) were also quite prevalent. Researchers have developed automated systems to ask [23] or answer [14] questions on social networking sites. Our findings indicate that blind users tend to ask more urgent and less subjective questions than the types of questions asked by sighted users in the aforementioned studies, which may require different types of socio-technical answering solutions.

Crowdsourcing information seeking using anonymous, paid workers such as those on Amazon's Mechanical Turk service [1] is an area of increasing interest. For example, the TailAnswers project [3] uses crowdsourcing to generate succinct inline answers (rather than sets of links) in response to search engine queries. The VizWiz project [5] employed crowdsourcing as an assistive tool for the visually impaired community, demonstrating an approach called quickturkit that allowed blind users to receive crowdsourced answers to their questions in nearly-real-time (less than one minute). [5] reported on a preliminary field deployment of VizWiz with 11 users. In this paper, we report on real-world findings from a year-long deployment of VizWiz Social, a tool based on the concept proposed in [5]; our dataset includes over 40,000 questions posed by over 5,000 users during this period.

VIZWIZ SOCIAL

VizWiz Social is a freely available iPhone application based on the crowd-sourced photo-based Q&A concept that Bigham et al. [5] tested with a pilot group of 11 blind users. As in Bigham et al.'s original application, VizWiz Social allows blind users to take a photo using their phone's camera, record a brief audio question (up to 15 seconds) to accompany the photo, and receive quick replies (on the order of a minute latency). Users' interaction is mediated by VoiceOver, a screenreader that comes pre-installed on the iPhone. Bigham et al.'s original application sent all questions to crowdworkers on Mechanical Turk; VizWiz Social offers users several options for answer sources, including crowdsourced workers from Mechanical Turk and/or VisionIQ [iqengines.com] (a computer vision service that uses human workers to manually identify unrecognized objects) and/or friendsourced workers (via email, Facebook, or Twitter messages to a user's contacts). Users choose answer sources after taking the photo and recording the audio. Although VizWiz Social expands on Bigham et al.'s initial concept by offering both crowd- and friend-based answers, the primary contribution of this paper lies not in the novelty of the VizWiz Social system, but rather in the understanding gained from a detailed analysis of users' interactions with the software.

VizWiz Social was released to the Apple App Store on May 31st, 2011 where it could be downloaded and installed by iPhone users for free. Users are not charged for asking questions via VizWiz Social; all costs associated with crowdsourcing are paid by our research group. In the first year that the application was available (May 31st, 2011 to May 31st, 2012) it was used by 5,329 users who asked 40,748 questions combined.

Users are identified by anonymous phone hardware ID numbers throughout the study, and no demographic information is collected by the application. However, a small previous study provided demographics for 12 active VizWiz users (4+ questions/month) - 7 were male; 4 were age 20-29, 6 were 30-39, and 2 were 50-59; 9 used Facebook and 8 used Twitter [7].

QUESTIONS ASKED

The questions asked by VizWiz Social users represent an incredibly diverse selection of accessibility issues encountered in everyday life. Questions helped users complete daily tasks (e.g., getting dressed, cooking meals) and can provide information about rare events (e.g., a child's illness, a mouse in the kitchen). By examining the types and features of VizWiz Social questions, we can identify what tools might increase blind people's independence, and gain insight into how to automate the answering process.

A random sample of 1000 questions was selected from the database for analysis. Questions were only selected from the 74% of users who agreed to a disclaimer allowing their questions and photographs to be used for research. Questions were categorized by type, primary subject, perceived urgency, subjectivity, and photograph quality (with specific methodologies presented in each section).

Question Types

The types of questions asked by VizWiz Social users provides an important first look at the accessibility challenges faced by blind people.

Methodology

Question categories, presented in Figure 1, were developed by a team of researchers on a different random sample of 1000 questions. Categories were developed through a two-stage affinity diagramming process [4]. The question text and images were presented simultaneously, and four researchers simultaneously did a silent, first-pass categorization of the queries. Once the researchers each developed their initial categories, they were merged and refined in a collaborative second-pass categorization and given names and definitions, resulting in the taxonomy in Figure 1.

Identification questions are those in which a user asks for an object to be identified by name or type (common subcategories include general identification inquiries lacking context, as well as those that are more specific such as identifying medicine, currency, or media). *Reading* questions are those in which the user requests that text be transcribed (common subcategories include mail, digital displays, numbers, hygiene-related information, and cooking-related information). *Description* questions are those in which the user requests a description of visual or physical properties of a depicted object (common subcategories include requesting descriptions of appearance, color, clothing, state settings, and

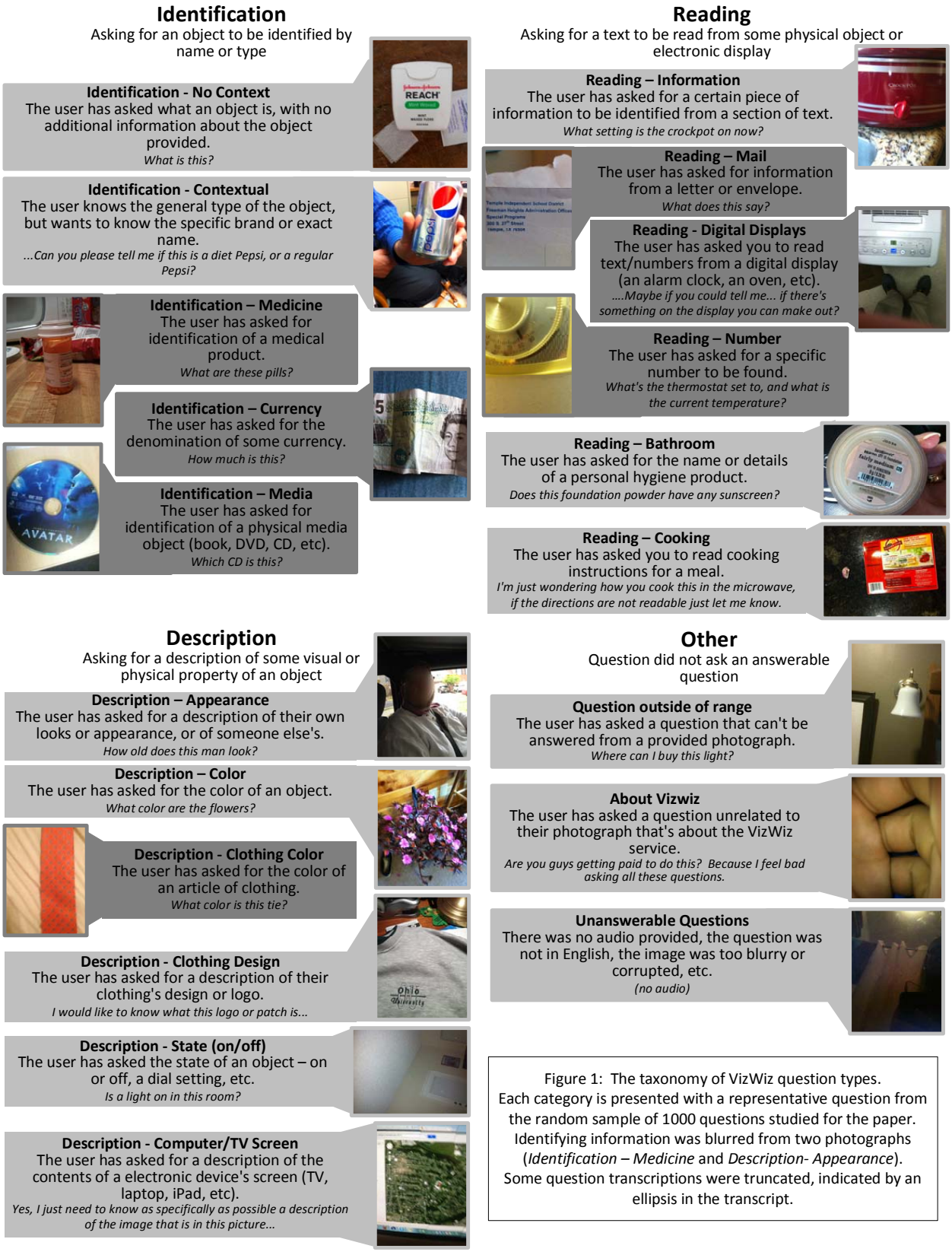

Figure 1: The taxonomy of VizWiz question types. Each category is presented with a representative question from the random sample of 1000 questions studied for the paper. Identifying information was blurred from two photographs (*Identification – Medicine* and *Description- Appearance*). Some question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

Figure 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.

digital screens). An *Other* category is used for unanswerable questions, such as those with unusable images, those that cannot be answered from the content shown in the photo, or those in a foreign language. Figure 1 provides examples of photographs and questions for each question type category and sub-category.

To apply this classification scheme to the target sample of 1000 questions (different than the ones used to develop the scheme), each question was examined in full with both the user's audio clip and photograph. Questions were then placed into one of the four over-arching categories (*Identification*, *Description*, *Reading*, or *Other*), and then placed into further sub-categories until they reached the deepest level that accurately described the question.

A second rater redundantly classified a random sub-sample of 50 questions into the four major category types, using a definition chart similar to Figure 1 to guide classification. The resulting Cohen's Kappa score of 0.425 indicates a moderate level of agreement.

Results of Question Type Categorization

Percentages of each category's representation in the sample are given below. Percentages represent the total percent of questions that were in a category or any of its subcategories, and have been rounded to whole numbers.

Identification questions were most common, making up 41% of the total sample. More than half of *Identification* questions were simple, "No Context" questions (58%) where the user did not provide any information about the subject of the photograph. Over a quarter of identification questions were slightly more complex "Contextual" questions (32%) where the user provided some starting knowledge about what the object was, but wanted a more specific identification (e.g., a brand name). The remainder of the questions were also given some context in the user's question, and featured either media (6%), medicine (2%), or currency (1%).

About a quarter of questions asked were *Description* questions (24%). 38% of the description questions were related to a user's outward appearance or dress - 24% asked about clothing color, 7% asked about clothing pattern or design, and 7% asked about physical appearance. 24% asked for color of other objects, 16% for the displays of computer or TV screens, and 2% for the physical state (e.g., on/off) of an object in the room. The remaining 18% of questions asked for general descriptive properties.

17% of the questions were *Reading* questions. Of the reading questions, 64% were seeking a specific subset of available textual content - 11% asked for numerical information, 6% for information from digital displays, 6% for information from letters or envelopes, and the remaining 41% for other readable information. 6% of questions asked for cooking instructions from prepared meals, and 3% asked for written information from bathroom objects such as shampoo or hairspray. Over a quarter were just general reading questions requesting an entire passage of content (27%), often of the format "Can you read this for me?".

The 17% of questions classified as *Other* were mostly due to audio issues - if a user didn't ask a question, the question wasn't in English, or the recording started too late or stopped too early (93% of *Other* questions). This may indicate the need to clarify the instructions on how to use the application, and give more explicit feedback on recording start time and length. We have recently added a tutorial to the VizWiz Social website (`http://www.vizwiz.org`) in order to give users more information on how to use the application correctly. The remaining *Other* questions asked questions that couldn't be answered from a photograph and were therefore out of the range of the service (6%), or for information about how the VizWiz Social application worked (1%).

Primary Subject of Photographs

In addition to analyzing the types of questions asked by users, we also examined the photographs separate from the audio questions that accompanied them. This allowed us to group questions based on their subject matter (rather than on their intent, as in the prior section) and offers insight into what categories of visual information VizWiz users could not easily access.

Methodology

Primary subjects in each of the 1000 photographs in our sample were first identified by a researcher. For each photograph, the researcher examined the photograph to determine what the subject was intended to be without having listened to the question. Photographs could also be marked as *Erroneous* if the photograph was blurry or too dark to identify anything, or *Unclear* if the photograph's quality was acceptable but no primary subject could be identified. The primary subjects were then grouped into categories and subcategories; this classification scheme was validated by another rater who used descriptions of each category (such as those in Table 1) to classify 50 randomly chosen images from our sample, resulting in a Cohen's Kappa of 0.516.

Categories of Primary Subjects

The major categories identified during the analysis of the primary subjects were: *Object*, *Setting*, and *Person/Animal*. *Object* encompassed any photographs featuring commercial product, household furniture, or other physical article as the primary subject. *Setting* encompassed any photographs which showed a whole room or an outdoor location as the primary subject. *Person/Animal* encompassed any photographs which showed a single person, a group or people or audience, or a pet as the primary subject.

In addition to these major classification categories, each response in the *Object* category was given a sub-category based on the type of object it was. Once all of the objects were classified, only the sub-categories with at least 5% of the objects were retained and the rest of the objects were instead added into a larger category, *Miscellaneous Objects*. Subcategory descriptions and percentages for the *Object* category are listed in Table 1.

Sub-category Name	Description	Percent
Food/Drink	A food product or beverage, either packaged or unpackaged	28%
Computer/TV	The screen of a computer or television, and any accessories (eg. 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 (eg. video game, CD, book)	6%
Paper	A letter or piece of paper	6%
Bathroom	Shampoo, conditioner, or other beauty and hygiene products	6%
Miscellaneous Objects	Any object that does not fit into another sub-category	12%

Table 1. Primary subject sub-categories and definitions for photographs in the *Object* category. This list does not include photographs with unclear primary subjects, or photographs that weren't usable

Results of Primary Subject Identification

Primary subjects were identified in the majority of the photographs (88%). Most of the photographs focused on members of the *Object* category (76%), with far fewer focused on members of the *Person/Animal* category (5%) and *Setting* category (4%). Within the *Object* category, most of the primary subjects fell into the *Food/Drink* category. In addition to the pictures where a primary subject could be identified, 6% were usable photographs where no primary subject could be identified, and 7% were photographs that were too dark, blurry, or out-of-focus to analyze.

Question Urgency

One aspect of mobile communication is that it enables users to get answers to *urgent* (i.e., time-sensitive) questions. By utilizing cellular phones, users can access resources while on the go and find answers to their information needs quickly. For blind users of VizWiz, web resources and family or friends are augmented by the constant availability of Mechanical Turk workers to answer questions.

We examined the perceived urgency of the questions asked by VizWiz users to find out if users were asking urgent questions via VizWiz. Urgent questions (such as a woman asking about a sick child, "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?") were questions with high stakes that required fast answers for action to be taken; less urgent questions (such as a woman asking, "What color is this wall?" in her home) could be answered in longer periods of time without much impact.

Methodology

Since VizWiz users did not specify their own level of urgency for their question, we labeled the *perceived urgency* of each question based on the amount of time in which the questions could be answered. The 1,000 questions in our sample were ranked by a researcher on a 5-point Likert scale. The scale was described with the following 5 levels of urgency.

1. **Within a minute:** The question asked must be answered in 60 seconds or less.
2. **Within a few minutes:** The question asked must be answered in 1 to 10 minutes.
3. **Within an hour:** The question asked must be answered in 10 minutes to 1 hour.
4. **Within the day:** The question asked must be answered in 1 to 24 hours.
5. **At any time:** The question can be answered at any time.

Questions were not categorized for urgency if there were problems with the photograph or audio (if the user did not record a question, or if the photograph was too blurry, dark, or out-of-focus to determine its subject). A second rater redundantly rated 50 questions randomly sub-sampled from the initial 1,000, with a Cohen's Kappa of .27 (fair agreement); the reader should bear in mind that this lower inter-rater reliability likely reflects the fact that it may be difficult for someone other than the original question asker to assess a question's urgency at a fine-grained level of detail. Urgency ratings were based on the raters perception of question urgency, not on direct feedback from users, and do not consider external sources of urgency (eg. wanting an answer before leaving a shop) or the users patience.

Results of Urgency Classification

The majority of the questions were classified as needing answers quickly - either within a minute (10%) or within 1 to 10 minutes (58%). Almost all the remaining questions were marked as not being urgent, and could have been answered at any time (1%), with a few needing answers within an hour (2%) or within the day (4%). Additionally, 25% of the questions were marked as being unsuitable for judging urgency. 98% of these questions had either been marked as *Other* question types (i.e., not possible to answer) or had poor photograph quality or unclear subjects, preventing a decision from being made about the urgency of the question.

Question Subjectivity and Objectivity

Another aspect of the questions that we examined was the *perceived subjectivity or objectivity* of the questions asked by VizWiz users. Some types of questions, such as product identification or text reading, are objective and require only observations about the photograph provided. Other kinds of questions, such as those about personal appearance or room cleanliness, are more subjective and may require the answerer to form their own opinion.

Determining the number and types of questions asked that are objective or subjective will provide information about the importance of the human workers involved in the VizWiz Social process. Certain objective questions (such as "What does this microwave say?") may contain objects or text that can be identified automatically based on the photographs provided, while subjective questions (such as "Does this [outfit]

match?" or "Is this person attractive?") may require human reasoning to answer well. We identify ratings of the objectivity and subjectivity of questions asked by VizWiz Social users, and discuss what question categories and primary subjects are contained in those questions.

Methodology

For the purpose of these experiments, we defined *subjective* questions to be "questions that are meant to be answered with opinions," and *objective* questions as "questions that are meant to be answered with observations or facts."

A researcher ranked the subjectivity or objectivity of the question on a 5-point Likert scale. A scale was used rather than a strict binary classification as subjective or objective, due to the nuanced nature of questions and answers revealed by inspecting our data set; for instance, even "objective" questions like identifying the color of an object can incorporate subjectivity due to asker-based characteristics (e.g., poor lighting quality in the asker's photo) or answerer-based characteristics (e.g., expertise or interest - a design student might answer "olive green" while a lazier or less artistic answerer might simply state "green"). The scale was described with the following 5 levels of subjectivity.

1. **Very subjective:** The question is asking for only opinions.
2. **Somewhat subjective:** The question is asking for mostly opinions, but observations or facts could be appropriate as well.
3. **Neither subjective nor objective:** Good answers to the question could be either opinions OR observations or facts.
4. **Somewhat objective:** The question is asking for mostly observations or facts, but opinions could be appropriate as well.
5. **Very objective:** The question is asking for only observations or facts.

Questions were not categorized for subjectivity or objectivity if there were problems with the photograph or audio (if the user did not record a question, or if the photograph was too blurry, dark, or out-of-focus to determine its subject). Scores were validated by a second rater for a random sub-sample of 50 of the 1,000 questions, with a Cohen's Kappa of .431.

Results of Subjectivity and Objectivity Classification

Nearly all of the questions that could be categorized were found to be objective - 61% were ranked as *very objective* and 17% as *somewhat objective*. Only 4% of the questions were ranked as being subjective (1% *very subjective*, and 2% *somewhat subjective*), and 1% were marked as *neither subjective nor objective*.

Additionally, 18% of the questions were marked as being unsuitable for judging subjectivity or objectivity. 91% of these questions were members of the *Other - Audio* category, since nothing could be learned about questions with poor audio.

Photograph Quality

We measured the quality of each photograph taken in order to determine how easily the question could be answered (e.g., by technical solutions such as automated object recognition software). When photographs are sent in that are blurry or out of focus, answers can be harder to find. However, some questions are answerable even with low quality photographs, due to questions that don't require much information (e.g., *Description - Color* questions can be answered from only a small section of the object, even if blurry) or questions where humans can make inferences (e.g., a question with a pink bottle where only "Pept" is visible is most likely Pepto Bismal).

Methodology

Each photograph was examined by a researcher and judged on its quality. The quality rankings went from 1 (worst) to 5 (best), and measured the quality based on several factors:

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?

Photographs were given an initial score of 5, and a point was deducted for each photographic error found. Photographs with a score of 1 corresponded to those marked as *Erroneous* in the earlier primary subject identification. Photograph quality ratings were based on human ratings rather than computerized analysis of features like blur and lighting levels in order to determine question suitability for human responders.

Result of Photograph Quality Ranking

Most of the photographs suffered from photographic errors. Only 18% of photographs scored perfect scores of 5. The majority of photographs had one or two photographic errors (33% scored 4, and 29% scored 3). Despite these errors, only 5% of the photographs with scores of 3 or 4 were marked as having *Unclear* subjects (based on our earlier photographic subject categorization). For the majority of photographs, it was possible for a human judge to determine what kind of object was visible in the picture, even if the full details might not have been shown.

13% of photographs scored a quality of 2. 23% of these photographs were marked as having an *Unclear* primary subject. As discussed in the primary subject identification section, 7% of photographs had a quality score of 1 and all were marked as *Erroneous* photographs. The average quality score for photographs was 3.41.

USER BEHAVIORS

In addition to examining a random sample of all questions sent to VizWiz Social, we also examined user encounters with the application. Our analysis of questions provided insight into the types of challenges blind users encounter in their day-to-day lives; our analysis of user-level behavior complements these findings by offering insight into the challenges and successes of blind users when adopting an access technology.

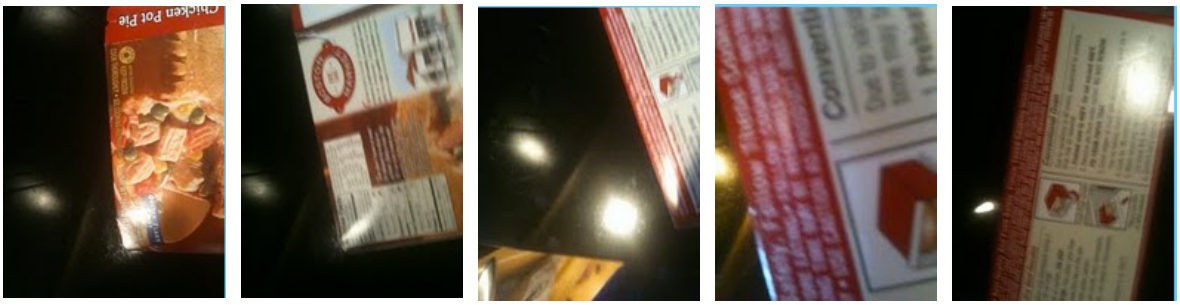

Figure 2. Sequence of pictures taken to answer, "How do I cook this?" with VizWiz Social. Each question was answered quickly, but cooking instructions took more than 10 minutes to receive as crowd workers helped the user frame the right part of the box.

We looked at a random set of users to determine how the application's "first impression" affected their continued use of the service. For those users who did continue to use the application, we analyzed how their question-asking behaviors continued or changed over time.

First Impression of VizWiz Social

A bad 'first impression' of an application - the inability to complete a task or slow response times - can impact a user's continued use of a service. For example, the creators of the Aardvark social search engine noted that "the strategy for increasing the knowledge base of Aardvark crucially involves creating a good experience for users so that they remain active and are inclined to invite their friends," and invested effort in creating a particularly good first-time interaction [15]. To examine the 'first impression' that VizWiz Social made on its users, we randomly selected 100 users and analyzed the kinds of questions they sent for their first interaction with the system, the answers they received, and whether they continued to use the application after their first attempt.

Return to the Service

We first looked at how many first-time users of VizWiz Social returned to it after one day. We classified a day as a 24-hour period from posting the first question. 45% of the first-time users in our sample used the service only for 1 day, while 55% used it for more than 1 day (and 50 of those used the service over multiple weeks, at least 7 days beyond their first use).

The average number of questions asked by a user was 10.3 (median 4). The single-day users asked an average of 2.56 questions. 24 of those single-day users asked only 1 question, with the remaining 23 asking an average of 3.92 questions in their single day of service. Multi-day users asked a much larger number of questions, with an average of 15.19 questions per user. The average number of days that all 100 sampled users used the service was 66.7. Among the 55 multi-day users, that number jumped to 120.45 days.

The First Question

The largest number of users asked "Identification" questions for their first question (44). 34 were simple, no-context identification questions (*"Identification - No Context"*) such as "What is this?". Often, these users asked this identification question for objects likely to be known (i.e., familiar objects posessing a distinct tactile shape, such as a cane), presumably testing the accuracy of the service. Of the remaining questions, 10 of the questions were "Reading" questions, and 19 were "Description" questions.

A number of users encountered difficulty when asking their first question. The number of questions which were categorized in the "Other" category was 27%, far exceeding the 17% of general questions which were in the "Other" category. Though 23 of the "Other" questions were due to the user not asking a question or asking in a foreign language (*"Other - Audio"*), 3 were questions that couldn't be answered from the provided photographs (*"Other - Range"*) and 1 was a question about using the VizWiz Social service (*"Other - VizWiz"*). The higher rate of *"Other - Audio"* questions for first questions (25% vs. 17% in the general sample) may indicate the learning curve that is present when users first use our application, while the *"Error - Range"* questions may indicate a user's confusion about the purpose of the VizWiz Social service.

18 of the 27 users (67%) who encountered errors while asking questions (i.e., photo of type "Other") did not continue using the service after their first day, including all 3 users who asked out-of-range questions. This is a significantly higher abandonment rate than for the 45% of overall users who abandoned the application after one day's use ($\chi^2(1, N=27) = 5.40$, $p = .02$).

Photography Quality

The average photograph quality for first questions was similar to the average general photograph quality (3.65 for first questions, 3.41 for general questions).

Answers

The majority of the questions sent in were answered correctly in the first answer received (for crowdsourced answers, a minimum of two answers was provided for quality control), if they were possible to answer. 78 questions received either a correct answer or, if in the *Other - Audio* category, a description of what was visible in the photograph. 9 users whose questions were unanswerable received feedback on how to improve their photographs.

However, 13 of the questions were answered incorrectly, presumably due to malicious or "lazy" turkers [2]. Some of these

questions received empty or nonsense responses, while others received answers that did not address their questions - for example, the first answer to the question "Does this shirt go with this skirt?" was simply "multi color clothe[s]." 9 of these 13 users (69%) did not ask any questions after their first day, and only 1 of them continued using the service for longer than a week. The negative first impression left by poor answers clearly impacted their continued use of the service.

Power Users

In addition to studying randomly selected users, we also examined the users who had the most experience with VizWiz Social. These "power users" used VizWiz Social over multiple months for large numbers of questions. We wanted to see if the use of the service differed from the random sample of questions asked by all VizWiz users, and if the types of questions they asked or their photography skills changed over time.

Methodology

The 25 most active users of VizWiz Social were chosen to be analyzed. The 25 selected power users had a much higher volume of questions per day than the randomly selected users, asked an average of 283 questions in their lifespan with the service, which averaged to 295 days. For each user, we analyzed 5 questions chosen from their first week of using the service, and 5 questions from their most recent week of use. The questions were then categorized by types and ranked in photograph quality as discussed in the earlier Questions section.

Question Types

The majority of first questions by power users were *Identification* (73%) questions. The three other categories of questions were not highly represented - *Description* questions were 14%, *Other* questions were 9%, and *Reading* questions were 4%. About half of identification questions were no context (52%).

In contrast, the highest proportion of recent questions by power users were *Reading* (46%) questions. *Identification* questions made up the next 25%, and *Other* questions were 21%. Only a small proportion of the questions were *Description* questions (8%) .

Photography Quality

The average photograph quality was significantly different between recent and old questions. Recent questions had an average ranking of 3.62 (SD = 0.998), while old questions had an average ranking of 3.32 (SD = 1.293). We performed the Aligned Rank Transform procedure [30] with photograph quality as a non-parametric measure, which revealed a trend level effect of time on quality, $F(1,216) = 3.15$, $p < 0.10$. There was also a significant interaction effect of time by photograph number on quality, $F(4,216) = 2.45$, $p < 0.05$. Raters were not told if photos were recent or old, and made their ratings blindly.

DISCUSSION

By analyzing 1,000 questions from the over 40,000 submitted by VizWiz Social users from May 2011 through May 2012, we gained an increased understanding of the challenges encountered in blind users' daily lives.

Users' inquiries were quite varied, mostly seeking objective information (such as transcriptions of text or descriptions of control states displayed on inaccessibly-designed appliances and gadgets); only a small portion requested subjective opinions such as aesthetic descriptions of users' appearance (which is a common use of Q&A services by sighted users [22]). Not surprisingly, interacting with non-accessible gadgets and digital displays was a source of many users' help needs; the prevalence of questions concerning food and cooking was also quite striking, revealing a more mundane (but quite important!) aspect of daily life with which blind people struggle to cope independently.

While 68% of questions were judged to be urgent by our raters, those ratings had low inter-rater reliability and we do not know how urgent the VizWiz users themselves considered their questions. In the future, more accurate question urgency levels could be obtained by asking VizWiz users to choose a maximum time they would be willing to wait for an answer.

Objective challenges might be eventually answered well by automated, vision-based algorithms, while crowd- and friend-sourcing are likely more appropriate for subjective inquiries. Friend-sourcing, whose efficacy is impacted by network size [21], may be ill-suited for addressing many of blind users' needs, however, due to the time-sensitive nature exhibited by most questions in our sample. Our analysis of photographic quality suggests that before automated approaches can be used for objective questions, they will have to contend with errors in blur, lighting, framing, and composition – such errors might be addressed by advances in computer vision technology, by replacing still-photo capture with a more interactive medium such as video and real-time communication, and/or by developing solutions that simplify blind photography [16]. We hope to perform further analysis of the photos taken by VizWiz users in order to determine the issues with photo composition and the differences in photo quality judgments made by humans and computer vision software, and to examine how this may impact the selection of the best answer source for a particular question.

Even though blind people who have an iPhone and installed VizWiz Social are likely to be a fairly self-selected group, willing to experiment with novel technologies, we saw that users who had a poor first experience with our application (either because their input to the system was of poor quality or because the system's answers were of low quality) had a higher-than-typical abandonment rate. This demonstrates that usability, in addition to utility, plays a key role in the adoption of access technologies. In future work, we hope to improve both the ease of photography for users and the quality of answers from the system to encourage user retention and allow us to study more users long-term.

Throughout the course of the year-long study, we also observed informally the effects of VizWiz Social users on one another. For instance, we observed VizWiz Social users sharing examples of types of questions that they found worked

well on the service via messages on Twitter. As one example, at the start of the National Federation of the Blind annual convention, a VizWiz Social user shared how helpful VizWiz Social was in identifying which bottle was which in his hotel room, e.g., shampoo, conditioner, lotion. For the next few days, and even today, we received a large number of questions of this type, forming a type of *meme* among users.

At higher level, our results clearly show that blind people have many of the same questions and concerns in their everyday lives as everyone else. They not only want to know about the relatively dry material that most technology targets, like reading mail and matching clothes, but want to know subjective information like whether their outfits look cool or appropriate for work, what the sunset looks like, and, yes, even the attractiveness of the people around them. To many working in the accessibility field this may not come as a surprise, but it is well-documented that everyday issues like social perception are too often ignored in the design of access technologies [26].

Most of the technology available today to help blind people interpret their visual world works much as VizWiz Social does – one picture at a time – and so understanding what users would want to do with this sort of interaction is important. However, as we move forward, a goal will be to expand the kinds of questions that VizWiz Social can answer with workers connected synchronously to users via streaming video so they can help for the duration of continuous activities.

CONCLUSION

In this paper, we reported on blind users' interactions with VizWiz Social, an iPhone application that supports photographic Q&A powered by crowd- and friend-sourcing. Analyzing data from "in the wild" use of this system by thousands of users over a year-long period, we gained insight into the challenges blind users sought assistance with, including analyses of question type, photo subject, question urgency, question objectivity, and photograph quality. We also examined adoption and usability issues specific to the VizWiz Social application, such as abandonment rates, first encounters with the technology, and behaviors of more experienced users. We hope that our findings provide inspiration for designing technical and socio-technical solutions to some of the many challenges encountered by blind users.

ACKNOWLEDGMENTS

We thank Jacob Wobbrock for his assistance in data analysis, and Kyle Murray for his assistance with affinity diagramming. This work has been supported by National Science Foundation Awards #IIS-1149709 and #IIS-1049080.

REFERENCES

1. Amazon Mechanical Turk, 2012. **http://www.mturk.com**.
2. Bernstein, M., Little, G., and Miller, R. Soylent: A word processor with a crowd inside. In *Proceedings of UIST 2010* (2010).
3. Bernstein, M., Teevan, J., Dumais, S., Liebling, D., and Horvitz, E. Direct answers for search queries in the long tail. In *Proceedings of CHI 2012* (2012).
4. Beyer, H., and Holtzblatt, K. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann, September 1997.
5. Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S., and Yeh, T. Vizwiz: Nearly real-time answers to visual questions. In *Proceedings of UIST 2010* (2010).
6. Bigham, J. P., Ladner, R. E., and Borodin, Y. The design of human-powered access technology. In *Proceedings of ASSETS 2011* (2011), 3–10.
7. Brady, E., Zhong, Y., Morris, M. R., and Bigham, J. P. Investigating the appropriateness of social network question asking as a resource for blind users. In *Proceedings of CSCW 2013* (2013).
8. ChaCha, 2012. **http://www.chacha.com**.
9. Chen, C., and Raman, T. Announcing eyes-free shell for android, 2009. **http://google-opensource.blogspot.com/2009/04/announcing-eyes-free-shell-for-android.html**.
10. Church, K., and Oliver, N. Understanding mobile web and mobile search use in todays dynamic mobile landscape. In *Proceedings of MobileHCI 2011* (2011).
11. Church, K., and Smyth, B. Understanding the intent behind mobile information needs. In *Proceedings of IUI 2009* (2009).
12. Google Goggles, 2012. **http://www.google.com/mobile/goggles**.
13. GoogleSMS, 2012. **http://www.google.com/mobile/sms**.
14. Hecht, B., Teevan, J., Morris, M., , and Liebling, D. Searchbuddies: Bringing search engines into the conversation. In *Proceedings of ICWSM 2012* (2012).
15. Horowitz, D., and Kamvar, S. D. The anatomy of a large-scale social search engine. In *Proceedings of WWW 2010* (2010).
16. Jayant, C., Ji, H., White, S., and Bigham, J. P. Supporting blind photography. In *Proceedings of ASSETS 2011* (2011).
17. Kane, S. K., Bigham, J. P., and Wobbrock, J. O. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of ASSETS 2008* (2008), 73–80.
18. Kane, S. K., Jayant, C., Wobbrock, J. O., and Ladner, R. E. Freedom to roam: A study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of ASSETS 2009* (2009).
19. KGB Answers, 2012. **http://www.kgbanswers.com**.
20. Lee, U., Kang, H., Yi, E., Yi, M., , and Kantola, J. Understanding mobile q&a usage: An exploratory study. In *Proceedings of CHI 2012* (2012).
21. Morris, M., Teevan, J., and Panovich, K. A comparison of information seeking using searchengines and social networks. In *Proceedings of ICWSM 2010* (2010).
22. Morris, M., Teevan, J., and Panovich, K. What do people ask their social networks, and why? a survey study of status message q&a behavior. In *Proceedings of CHI 2010* (2010).
23. Nichols, J., and Kang, J.-H. Asking questions of targeted strangers on social networks. In *Proceedings of CSCW 2012* (2012).
24. Paul, S., Hong, L., and Chi, E. Is twitter a good place for asking questions? a characterization study. In *Proceedings of ICWSM 2011* (2011).
25. Red Laser, 2012. **http://www.redlaser.com**.
26. Shinohara, K., and Wobbrock, J. O. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of CHI 2011* (2011), 705–714.
27. Smith, A. Americans and their cell phones. Pew Internet American Life Project, 2011. **http://pewinternet.org/Reports/2011/Cell-Phones.aspx**.
28. Sohn, T., Li, K. A., Griswold, W. G., and Hollan, J. D. A diary study of mobile information needs. In *Proceedings of CHI 2008* (2008).
29. Voiceover, 2009. **http://www.apple.com/accessibility/voiceover/**.
30. Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of CHI 2011* (2011).