

# Accessible Photo Album: Enhancing the Photo Sharing Experience for People with Visual Impairment

Susumu Harada<sup>†</sup> Daisuke Sato<sup>†</sup>  
Hironobu Takagi<sup>†</sup>

<sup>†</sup>IBM Research - Tokyo  
NBF Toyosu Canal Front, Toyosu 5-6-52, Koto,  
Tokyo, Japan  
{haradas, dsato, takagih, chie}@jp.ibm.com

Dustin Adams<sup>‡</sup> Sri Kurniawan<sup>‡</sup>  
Chieko Asakawa<sup>†</sup>

<sup>‡</sup>University of California, Santa Cruz  
1156 High Street, Santa Cruz, CA 95064, USA

{dustinadams, srikur}@soe.ucsc.edu

## ABSTRACT

While a photograph is a visual artifact, studies reveal that a number of people with visual impairments are also interested in being able to share their memories and experiences with their sighted counterparts in the form of a photograph. We conducted an online survey to better understand the challenges faced by people with visual impairments in sharing and organizing photos, and reviewed existing tools and their limitations. Based on our analysis, we developed an accessible mobile application that enables a visually impaired user to capture photos along with audio recordings for the ambient sound and memo description and to browse through them eyes-free. Five visually impaired participants took part in a study in which they used our app to take photographs in naturalistic settings and to share them later with a sighted viewer. The participants were able to use our app to identify each photograph on their own during the photo sharing session, and reported high satisfaction in having been able to take the initiative during the process.

## Author Keywords

Visual impairment; blind; photo sharing; audiophotography.

## ACM Classification Keywords

H.5.0 [Information Interfaces and Presentation]: General.

## General Terms

Human Factors; Design; Measurement.

## INTRODUCTION

People take photographs for a variety of reasons, but one common reason is to visually preserve an impression of a particular experience – be it the company of loved ones, a beautiful view, or an object that evoked some distinctive emotion – to later revisit and reminisce upon or to share with others and recount the captured experience.

A survey conducted by Jayant et al. [8] revealed that many people with visual impairment<sup>1</sup> also share an interest in photography for much of the same reasons as those who are sighted; namely, to capture photographs of various moments and experiences to share with their sighted friends and family. The proliferation of social networking services that facilitate online sharing of digital photos are also cited as the motivating factor. Many of them are also actively engaged in or have tried taking photographs on their own or through the help of a sighted companion. There are also various courses offered that are tailored specifically to teach photography to people with visual impairment [16,17].

While the number of people with visual impairment who are engaging in photography is starting to increase, they still face a number of challenges. Aside from the technical challenge of capturing an “acceptable” photograph (typically regarding such aspects as framing, lighting and focus) on their own, there is the challenge of managing the captured photographs without being able to visually inspect them to distinguish one from another. There is also the challenge of recollecting the circumstance under which a photograph was taken without the benefit of the detailed visual cues that may be present in the photograph.

A number of researchers have been investigating methods to support blind people in taking “good” photos [8,13,14], but there have not been much investigation yet into how to effectively support blind users in managing, browsing through, and sharing the photos they capture. Our survey of people with visual impairments with an interest in photography revealed that a number of people are using various methods for associating some form of description to captured photographs, but many of these methods are far from optimal end-to-end experiences for visually impaired users in capturing and sharing photographs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

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

Copyright © 2013 ACM 978-1-4503-1899-0/13/04...\$15.00.

---

<sup>1</sup> In this paper, we use the term “people with visual impairment” and “partially sighted” to refer to people who experience significant difficulty in visually inspecting and identifying the content of photographs.

A promising development has been the emergence of smartphones that not only offer high-quality camera and network connectivity but also accessibility features such as iOS's VoiceOver [18] that enable a blind user to use the various applications on the device eyes-free. The rich set of functionalities offered by such smartphones also enables developers to build various applications for it, such as those that enable audio to be recorded alongside a photo capture [10,11,15,19,20,21]. While these applications have mostly been designed for sighted users and thus have a number of usability issues for use by visually impaired users, we see a great potential for utilizing such *audiophotographs* as a medium for supporting photo sharing and management for the visually impaired users.

To better understand how the photo browsing and sharing experience can be enhanced for people with visual impairment, we first conducted an online survey targeting people who are blind or partially sighted to inquire about their photo-taking and sharing experiences. We then built an accessible smartphone application that enables the user to capture audio recordings along with each photograph and to peruse them at a later time. A user with visual impairment can use the audio recording feature to associate a spoken description with each photograph to aid in identifying the photo later, as well as to record any ambient sound. To evaluate the effectiveness of our application and to observe the actual practice of taking and sharing photos, we conducted a user evaluation with five blind and partially sighted participants in naturalistic settings.

Our observations revealed that our accessible user interface and the audio recordings were effective in enabling the participants to later locate and identify the photos to be shared with a sighted observer. We also identify a number of insights and design considerations for making the photo sharing experience more accessible and inclusive. Most importantly, the participants reported great enjoyment and satisfaction in being able to engage in the activity of capturing photos and sharing them with others.

#### **RELATED WORK**

A substantial set of references indicate that blind and partially sighted people do partake in photography [12,16,22,23,24,25]. A survey conducted by Jayant et al. [8] revealed that over 70% of the respondents expressed interest in being able to use a camera accurately, with roughly the same number of respondents having actually used a camera recently to take photographs. Of those who had used a camera, the most prevalent reason was to capture photographs of friends, family, trips and events.

As visual inspection of photos is not feasible for visually impaired persons, non-visual clues such as a description, date, time and location the photo was taken, need to be associated with each photo in some way. Current practices include a wide range of methods, from using a Dictaphone or a Braille memo taker to jot down descriptions of each photo in sync with the sequence of photos captured on the

camera, to renaming the filename of the digital photograph to some descriptive text. In particular, we see a great potential in the use of audio recording to augment photographs, since it not only enables easy recording of descriptive text, but can also be used to capture ambient sound that may provide contextual cues when later recalling the content of the photograph.

#### **Augmenting Photographs with Audio**

An *audiophotograph*, as presented by Frohlich and Tallyn [5,6], is a medium that combines a photograph with associated audio recording as a single unit. The type and method of audio capture can range from recording ambient audio the moment the photo was taken, to adding a voice commentary, voiceover or associated music afterwards. Although audiophotograph is not a medium that was devised specifically for people with visual impairments, it does provide a convenient mechanism for capturing, organizing, and sharing captured moments among both sighted and visually impaired people.

While it may seem that an audiophotograph is similar in nature to video, there are certain notable differences. We do not discount the utility of video as a capturing medium, but it does have certain properties that make it less ideal for visually impaired users. For instance, the nature of video capture makes it such that the user would have to hold the camera steadily pointing at the target of interest while s/he records the ambient sound or memo audio, a task that may be particularly difficult for a blind person compared to snapping a photo and then recording the audio without worrying about where the camera is pointing.

There are also various freely and commercially available applications for smartphones that support audiophotography, such as Picle [10], Com-Phone [11], Speature [15], StoryMark [19], Voicepic [20] and StoryRobe [21]. While these applications do provide non-visual cues for identifying photos, the steps required for completing simple tasks such as snapping a photo and retrieving a photo from the library are cumbersome, sometimes requiring as many as 35 gestures to snap one photo when using the screen reading feature of the smartphone.

#### **Photo-taking Support for People with Visual Impairment**

Smartphones are becoming more accessible to visually impaired people, with the introduction of features such as VoiceOver on the iPhone [18] and TalkBack on the Android platform, leading to adoption of smartphones by members of the blind community. Combined with the powerful processors and high quality cameras on such smartphones, many applications are being developed that utilize the smartphone camera to enable visually impaired users to "view" the environment around them. Applications as VizWiz [3] and IQ Engine's oMoby [26] let users snap photos and have their contents be automatically identified. The app first tries to identify the content of the photo using computer vision techniques, and if unsuccessful, sends the

photo to a crowdsourcing service. This allows a blind user to distill otherwise inaccessible visual information about their environment. VizWiz::LocateIt [4] utilizes a similar technique to interactively guide the user towards a specific object within a scene.

To support people with visual impairment to take photos, some basic technology such as detecting blur in photos [9] and sonifying visual images [27] had been reported to be useful. One technique that was reported utilized accelerometers inside mobile devices to adjust camera tilt during photo-taking. EasySnap [14] provides users with audio feedback about blur and darkness of the camera view, camera tilt, as well as location and size of people's faces. PortraitFramer [8] also utilizes similar feedback methods to support interactive exploration of the location and size of multiple faces within the camera view. These are promising developments towards providing more control to people with visual impairments in the photo capturing process.

### PHOTO SHARING SURVEY

To further investigate the current practices surrounding the sharing of photos and the challenges faced by people with visual impairment, we conducted our own online survey targeting people with visual impairment. The survey was hosted on Survey Monkey during the summer of 2012 and email and Facebook invitations were sent to various mailing lists for people with visual impairment. We acknowledge that as a result of recruiting our participants on-line, our results may suggest a disproportionately large number of people engaged in photography and photo sharing.

We received responses from 47 people with visual impairment, ranging in age from 17 to 76, with the median age of 48. Thirteen people were totally blind or had light perception only, and seven people were blind from birth.

Of the respondents, 74% had experience taking photographs on their own, 87% has had somebody else take a photo for them, and 70% had shared their photos online. When asked which online service they used to share their photos, 25 reported using Facebook, 18 reported using direct email, eight reported using Twitter, three reported using Photobucket, and three reported using LiveJournal.

Twenty-seven percent of the respondents who have shared their photos responded that they were somewhat or mostly satisfied with their current method of sharing, with 52% responding that they are somewhat or mostly dissatisfied. Forty-three percent of the respondents answered that they quite strongly or very strongly wish they could share their photos online. Fifty-seven percent of respondents said that they mostly or very much wish they could take their own photos, and 49% responding they would like to be able to share their photos online.

Our survey also contained open-ended questions. A quarter of our respondents cited the inability to judge the quality and the content of the photo as the main inhibitor preventing them from being able to organize or share their

photos. Examples of actual statements include, *"I would like to immediately be able to label photos so I don't forget what they are. This is my biggest problem, remembering what I took photos of and in what order."* Another respondent mentioned, *"I don't know which ones are good enough to share."* As expected from the previous survey [8], 16 respondents cited difficulty with the act of taking good-quality photos including issues about focusing and aiming. Four respondents mentioned that photo sharing web services are not fully accessible for them.

One common theme that came up is the desire to have an accessible way to identify the picture or share it, as in *"it would be great to confidently access time/date and description of photo or circumstance. If I could attach a verbal description to the shot, that would at least give me the feedback I need when sharing in person or online."* Another user commented, *"I would love an accessible way to share my photos with friends and family!"*

### ACCESSIBLE PHOTO ALBUM APPLICATION

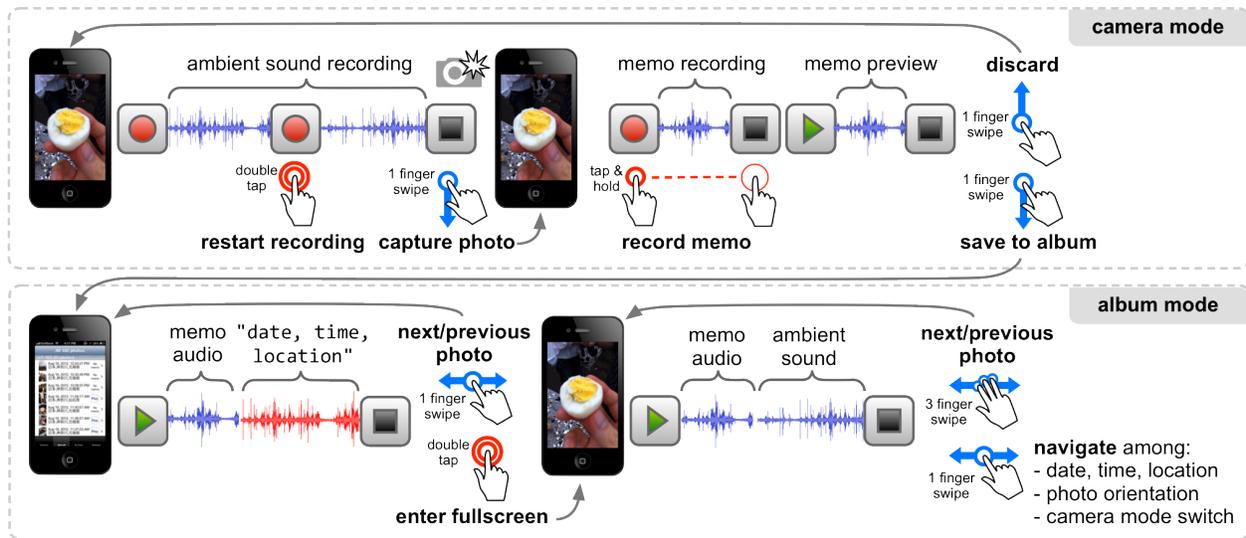
Through our literature survey and our own survey, we found that there are still unmet needs in supporting visually impaired persons to take, organize and share photos. To address these issues, we decided to create our own accessible application with the following design goals:

1. to enable the user to quickly capture a photograph along with the memo and ambient audio recordings,
2. to enable the user to quickly browse through the photo collection and grasp the content of each photo, and
3. to enable the user to take the lead during the activity of sharing their experience through their photos with a sighted person.

### Implementation

We developed an iPhone application that enables the user to capture a photo and augment it with an audio memo recording as well as a recording of the ambient sound. The app was designed to be accessible and optimized for operation via the built-in VoiceOver screen reader functionality of the iOS, so that a visually impaired user can use it without needing to see what is on the screen. The key functions were designed to be accessible through the various touch-screen gestures.

Figure 1 provides an overview of the functionality and interaction made available by our iPhone app. The app consists of two main modes: camera mode, and album mode. As soon as the app is launched, it enters the camera mode and ambient sound recording is started. The user can explicitly restart the ambient sound recording by double tapping on the screen. The user can capture a photo by performing a single-finger swipe-down gesture. While ideally we would have liked to utilize the physical volume button as a shutter, we encountered technical issues in implementing such a feature in conjunction with the audio recording capabilities, thereby resorting to using a gesture.



**Figure 1: Accessible Photo Album iPhone app overview showing the main functionalities in camera mode and album mode, designed to be used in conjunction with Apple iOS's VoiceOver screen reader functionality. In camera mode, the user can record ambient sound, capture a photo, and record a memo description. In album mode, standard iOS VoiceOver gestures are used to navigate among the photos. The memo audio is played each time a new photo is selected, and iOS VoiceOver speaks out supplemental information such as the date, time, location, and orientation of the photo.**

Most of our participants did not have significant issues with using the gesture-based camera shutter.

In order to record a memo audio, the user single-taps on the screen and keeps it held down for a second, at which point a confirmation sound is played and the microphone starts recording until the user releases the finger from the screen. At this point, the memo that was just recorded gets played back automatically for review, and the user can either re-record the memo by repeating the procedure, or accept and save it along with the photo by performing a single-finger swipe-down gesture again.

Under the photo-browsing mode, the user is first presented with a list view of photos sorted and grouped by days. The user can navigate from one photo to the next by using single-finger swipe-left or swipe-right gestures, as well as jump from one day to the next using swipe-up or swipe-down gestures. Each time a new photo entry is selected, its memo audio is played back, as well as the date, time and location of the photo (read out in human-understandable form, reverse-geocoded from the GPS location saved during photo capture) announced by VoiceOver's text-to-speech engine. When the user wishes to show the photo to a sighted person, the user can double tap on a specific entry in the list view, which displays the photo in full-screen view. In this view, the user is able to flip through each photo by using the VoiceOver-standard three-finger swipe-left and swipe-right gestures. Each time a new photo is displayed on the screen, the app automatically plays back the recorded memo audio, followed by the ambient audio recording. After the audio finishes playing, the orientation of the photo is also read out, such as "landscape, clockwise", indicating to the user that the photograph being shown was taken in landscape mode and thus the device

should be rotated clockwise from its standard portrait orientation when showing it to the sighted person.

The key advantage of our app is the tight integration of the audio recordings with the photographs in both the capturing phase and the browsing phase, and the optimization for operation in conjunction with the iOS VoiceOver functionality. By enabling the visually impaired user to flip through the photos while at the same time hearing the audio recordings, the user can quickly grasp which photo is which.

#### USER STUDY

To evaluate the effectiveness of our application and to observe the actual practice of taking and sharing photos by visually impaired users, we conducted a user study with blind and partially sighted participants who expressed interest in the idea of taking and sharing photos. More specifically, the goals of our user study were as follows:

1. To answer the following questions about whether our application was able to meet our goal:
  - (a) Did our app enable the user to quickly capture photographs along with memo and ambient audio recordings?
  - (b) Did our app enable the user to quickly browse through the photo collection and grasp the content of each photo?
  - (c) Did our app enable the user to engage proactively during the activity of sharing their experience through the photos with a sighted person?
2. To gain insights about future design considerations for making the photo sharing experience more accessible and inclusive, by observing visually impaired people

	Age	Gender	Visual impairment	Have taken photo on their own?	Have shared photos?
P1	25	Female	Left eye: hand-movement (since birth) Right eye: finger count (since age 15)	Yes (mobile phone camera)	Yes (show on phone, e-mail)
P2	34	Male	No light perception (since birth)	No	Yes, others' photos (e-mail)
P3	35	Male	No light perception (since birth)	No	No
P4	34	Female	Left eye: hand-movement (since age 4) Right eye: no light perception (since age 4)	Yes (mobile phone camera)	Yes, experimentally (e-mail, blog)
P5	36	Male	No light perception (since age 21)	Yes (mobile phone camera)	Yes (show on phone, e-mail)

**Table 1: Summary of background information about our user study participants.**

engaged in the end-to-end process of capturing, browsing through, and sharing their photographs.

Five participants (three males and two females ranging in age from 25 to 36) took part in our user study as paid volunteers, with each participant taking part in a full-day evaluation session. Information about each participant is summarized in Table 1, with details presented in the section following the study procedure presented next.

### Participants

Summary information about the five participants who took part in our study is presented in Table 1.

#### *Prior experiences with photo taking*

P1, P4, and P5 had previously used a mobile phone camera to capture photos, and had used the method of renaming the captured photo file with the memo text using the keypad-based text entry. P1 and P4 mentioned that they were relatively comfortable with the process, but P5 mentioned that it was too tedious for him and that he mainly relied on the timestamp of the captured photo to identify the captured photos, given that he only took a few photos. Both P2 and P3 had only tried taking photos on their own experimentally, but P2 had also had photos taken for him by his friends using his camera phone.

As expected, everyone expressed concerns about their ability to capture “good” photos on their own, especially of targets that are beyond reach such as a scenery or a group of people, as well as concerns over lighting, focus, and occlusion of the lens by their own fingers.

While the challenge of capturing photos independently was expressed by all participants, they also mentioned that they often simply had a sighted person take a photograph for them, or at least receive guidance in pointing the camera at an appropriate angle. During our pre-study interview, we also observed that for tangible objects within reach, P3 and P5 were particularly adept at aiming and positioning the camera appropriately. None of the participants had received any formal instruction on how to take photographs, so it is

also conceivable that their skills will increase given such opportunity.

#### *Prior experiences with photo sharing*

P1, P4, and P5 had experience sharing photos they had taken on their camera phone with others, by either showing the photos directly on their phone or by emailing them. P4 had tried uploading her photos to a blog a few times.

P2 had not taken his own photo, but for photos taken by his friends that he had wanted to share, he had uploaded them directly to a social networking site immediately after they were taken.

P3 expressed keen interest in wanting to take and share photos of his cooking on places like Twitter, and also mentioning that he’s come to believe that an impactful image is paramount when transmitting information to sighted users on places like Twitter and blogs.

### Procedure

Each of the five participants took part in a full-day user study session that consisted of four parts: an introductory session, in-the-wild photo-capturing session, in-lab photo-sharing session, and post-study interview session.

For each participant, we first had them come in to the lab for an introductory session, during which we interviewed them about their past experience with and attitude towards photo taking and sharing. We also introduced them to our application and demonstrated how to take a photograph and record the memo audio and the ambient audio, as well as how to browse through the collection of photos and audio recordings. We demonstrated with several example captures how the memo audio and the ambient audio are played back during the browsing mode, and how they could be used to determine the content of a particular photo as well as to share the ambience with the person to whom the photo is being shared. We described to them that they will next be escorted through a series of local tourist attractions, and that we would like them to think about sharing their experience later with a sighted person by capturing various contexts they find of interest as photographs using our app,

annotating each one with voice memo and ambient sound recordings as they see fit. The introductory session on average lasted about one and a half hours.

After the introductory session, an experimenter from our group escorted the participant through a series of local tourist attraction spots for the in-the-wild photo-capturing session. In an attempt to provide as naturalistic a setting as possible in which the participant would find compelling reasons to capture photographs, we had asked the participants prior to the study for their interests and places they've been wanting to explore. We then prepared customized travel routes for each participant, attempting to ensure that there will be at least some opportunity to encounter tangible objects of interest or something they can directly experience, instead of simply a series of "picturesque views." The experimenter provided mobility assistance for the participants as well as having casual conversations and describing the surrounding environment as they toured the various spots. While there was a pre-determined sequence of spots for each participant, we let the participant decide when to move on, at times letting them deviate from the original course if they found something else of interest, as well as letting them decide when to take a photograph. The participants were also free to ask the experimenter to take a photo of themselves in front of some scene, in which case the experimenter captured the photograph but let the participant record their own memo audio. The experimenter also assisted the participant whenever they requested help in pointing the camera appropriately in some desired fashion, by offering verbal instructions first before resorting to direct manipulation of the camera orientation. This is because we wanted to give the participants as much control over the capture of their own photos, and only take physical control of the camera when explicitly asked to so by the participants.

After spending approximately five hours out in the field, the participants were escorted back to the lab for the in-lab photo sharing session, where they were first asked to respond to a simple questionnaire about their experience of taking photos. They were then introduced to another member of our lab who served as the "viewer" to whom the participant would share their photos they had taken. The participants were simply asked to share their experiences throughout the day with the viewer by showing the photographs that were taken and describing them. Since some participants expressed hesitation about having their voice memo be heard by someone else, the participants were given the option of using an earphone and only letting the viewer hear the sounds whenever the participants so desired. The viewer was instructed to have a normal conversation with the participant, reacting to the photos and asking any questions about the photo or the participant's experience. The experimenter who accompanied the participant during the photo-capturing session also sat in on the photo sharing session, but played the role of a silent

observer. The viewer was also instructed to wait until the participant started to describe each photo on their own before making any comments or asking any questions about it. This was to first observe whether the participant was able to identify the photo based solely on their own memo recording, without receiving help from sighted people. If the participant explicitly expressed uncertainty about the content of the photo, the experimenter was allowed to naturally describe what she sees in the photo. The photo sharing session lasted approximately 30 minutes.

The full-day session was concluded with a post-study interview session during which we asked the participants for their comments and reflections about their experience throughout the day.

## Results

We first summarize the results from our study in the context of the application design goals. Given that the focus of our study was more towards observing the end-to-end user experience of photo capturing and sharing rather than on conducting controlled lab experiments, we have chosen to evaluate whether or not our goals were met based on the responses by our participants to our qualitative questionnaire. We acknowledge the importance of further validating our results via quantitative experiments, and we discuss ideas for such experiments in the Discussion section later in the paper.

### *Did our app enable the user to quickly capture photographs along with memo and ambient audio recordings?*

While only P3 had prior experience with the VoiceOver functionality of iOS mobile devices, all participants were able to grasp the basic gestures involved in using our app within the introductory session.

Each participant captured between 13 and 20 photos during the photo-capturing session over the period of approximately five hours (Table 2). In response to the statement "taking photos was easy" to which the participants responded on a 7-point Likert-scale with 1 being "strongly disagree" and 7 being "strongly agree," two participants responded with a 7, and the others responded

	<b>Photo-capturing session locations</b>	<b>Photos captured</b>
<b>P1</b>	Major waterfront leisure area with shopping malls and attractions	17
<b>P2</b>	Major landmark tower and major waterfront leisure area	18
<b>P3</b>	Wine import store and major waterfront leisure area	20
<b>P4</b>	Historic site and adjacent park	13
<b>P5</b>	Chinatown and uptown shopping mall	15

**Table 2: Summary of the places toured during the photo-capturing session and the number of photos captured.**

	P1	P2	P3	P4	P5
<b>Photo-capturing session</b>					
Taking photos was enjoyable	7	4	6	7	6
Taking photos was easy	5	7	5	7	5
Recording memo was easy	6	7	6	7	5
Felt greater desire to become able to take photos independently	3	4	7	7	4
<b>Compared to previous methods</b>					
Assigning memo to photo felt easier	7	-	-	6	7
Assigning memo to photo felt faster	7	-	-	6	5
Able to record more detailed memo	6	-	-	7	6
<b>Photo-sharing session</b>					
Browsing through photos was easy	7	7	6	7	7
Was able to identify each photo	6	5	7	7	7
Was able to recall each photo's context	7	6	6	6	7
Enjoyed being able to show my photos	7	7	7	6	6
Was able to play a proactive role during the photo sharing session	6	4	6	6	7
<b>Overall session</b>					
The experience of sharing the photos was enjoyable	7	6	6	7	6
Feel the desire to use an app like this in the future to share photos with others	6	7	6	7	3
Feel greater desire to try capturing photos than before the study	6	6	6	7	7
Feel greater desire to try sharing photos with others than before the study	7	7	6	7	4

**Table 3: Responses to the questionnaire regarding the photo capturing session, the photo-sharing session, and the overall session. Responses were on a 7-point Likert scale (1: strongly disagree, 2: disagree, 3: slightly disagree, 4: neither agree nor disagree, 5: slightly agree, 6: agree, 7: strongly agree).**

with a 5 (Table 3). Responses to the statement “recording memo was easy” included two 7s, two 6s and a 5.

For the three participants who had previous experience with capturing photos on their camera phones (P1, P4, P5), we also asked them to rate their agreement with the statements comparing the ease, speed, and level of detail of the memo recording using our app versus their previous methods. Across all three questions, all participants rated our app as being easier and faster in assigning memos to photos, and as enabling recording of more detailed memos.

*Did our app enable the user to quickly browse through the photo collection and grasp the content of each photo?*

All the participants quickly familiarized themselves with the basic gestures for flipping through the full-screen photos, controlling the audio playback, and determining the correct orientation of the photograph. All five participants were able to recall and identify the content of every photo, except for two photos that were unable to be distinguished by one of the participants (P2). The photo that led to the ambiguity was a photo of the participant captured by the experimenter, taken from knee height pointing upwards with the landmark tower in the background soaring behind

the participant. The memo description that the participant had recorded for the photo was “tried taking a photo of the landmark tower looking directly down upon it.” Upon hearing the memo during the photo sharing session, P2 became confused, given that the preceding photo was also of the landmark tower but one which the participant himself had taken, with the memo description “tried taking the landmark tower. Unfortunately we couldn’t go up to the observation deck, so looking up at it from below.” Aside from this one instance, all participants were able to advance through the photos using the three-finger flick gesture and after hearing the few seconds of the recorded audio memo for a particular photograph, begin recounting its context without any prompt from the viewer.

*Did our app enable the user to engage proactively during the activity of sharing their experience through the photos with a sighted person?*

In response to the questionnaire statement “felt able to play a proactive role during the photo sharing session,” all but one participant (P2) replied agree or strongly agree (Table 3). P2 described the reason why he felt neutral about this statement as being due to the one pair of photographs that he was unable to successfully distinguish, as mentioned above. While it is difficult to gauge how much that particular experience may have skewed his assessment of the experience over the entire set of photos, based on our observation, he was able to accurately recount the contents of all other photos without any problem.

### Observations

In this section, we present other notable observations that we made during the course of our user study.

#### *Ambient Audio Helped With Recall of Additional Details*

There were several instances in which the ambient audio served as a cue for the participant or the viewer to notice additional aspect of the captured context. When P5 was showing a photo he took of a Chinatown gate, the ambient audio started playing and he heard what sounded like water being sprayed. “Oh wow, I didn’t even notice someone was spraying water right next to me!” he commented.

P1 and P4 both initially chose not to let the viewer hear the recorded audio during the photo sharing session, citing embarrassment over having their memo recording be heard, although P4 did let the viewer hear just the ambient audio for two of the photographs (one of a fountain and another of a brass band). At the end of their sessions, however, they both commented on the value of the audio recordings.

*I figured that the ambient recording could capture not just random sounds but also the person I was with describing various things to me, so I could just record that and then take the photo and I can get all the information I need. – P1*

*I actually think the ambient audio is kind of nice. Not just when I’m listening to it, but I could see how it can convey more to even a sighted person if they could also hear the*

sounds. Also most of my friends are blind, so it's also nice that I can share those sounds with them as well. – P4

#### **Positive Comments About the Overall experience**

There were many positive comments about the photo sharing experience.

*The thing that became clear to me with this app, is that by being able to record my memos along with the photos, I can now manage my own photos. This is a huge step, a new level for me. Up until now, it was really a shot in the dark, trying to manage my photos based on very limited information. – P3*

In response to the question about how enjoyable the overall photo sharing experience was, one participant expressed her concern regarding the quality of her photos but also her excitement of being able to share the photos.

*It was really enjoyable. But I think I'll rate it a 6, because I was a bit unconfident about how well the photos came out, whether the sighted person would enjoy it, how much will be conveyed. But aside from that, I really liked the fact that I can explain to the other person that I actually went to these places, not just with words but also with photographs. So far I really haven't had such experiences, I've just been explaining to others about the experiences I've had just through words. But to the sighted, when I can explain with photographs, with just a glance they can see, ah, it was like that. And then we can have a conversation about it. When they comment about the contents of the photo, it's also like a feedback to myself, I can reconfirm what I've experienced. It was really fun. – P4*

Another participant commented on the added value gained from the ambient audio in being able to recall the situation.

*It's really great that I can tell right away what photo is being displayed, and when I hear the voice of the person I was with in the recording, it's like the things that are not in the photograph were also captured. It makes me feel like I got something extra for free. Like that lady speaking Chinese, or that other person shouting, "wait for me!" You just can't capture that in a photograph. – P5*

#### **INSIGHTS AND DESIGN CONSIDERATIONS**

Based on the results from our user study, we present the following key insights and design considerations for expanding the expressive options for people with visually impairment and in identifying new ways in which sighted and non-sighted people can share their experiences.

##### **Ability to Take the Lead During Photo Sharing**

One of the key design goals in creating our accessible photo album app was to provide a way for visually impaired users to be able to take the lead in sharing their experience through their own photographs. Although P5 had mentioned that he is currently able to identify his photos on his camera phone solely based on the timestamp because he hasn't captured very many photos, it is easy to imagine that as the

opportunity to capture photographs increase, identifying them will become increasingly challenging.

For photographs that one has captured on their own, it is possible to associate with it one's own memory about the context, the situation, and the reasons that led to the moment being captured. But if one has to rely on the help from a sighted person in retrieving that mapping each time, the enjoyment of photo sharing may be diminished.

P3 commented that he wishes he was able to rearrange the photos according to his desired order so that he could tell "his" story. During the photo sharing session, for a particular photo he temporarily attached the earphone to the device so that only he could hear the memo, flipped the device over so that the viewer could not see the screen, waited until he confirmed the memo, built up to a climax by saying "and now you won't believe what kind of bottle I came across next," and then finally flipped the phone over to reveal the photograph of an elaborately shaped bottle of whiskey. The aim of our research is not only to make the process of photo taking and sharing more accessible, but to explore ways in which we could expand the expressive options for people with visual impairment in such ways.

##### **User Interface Design Considerations**

There are many possible designs for the user interface that support capturing of audio along with photographs. Some key considerations are 1) in what order to capture the audio and photo, 2) how to display them, and 3) what the interaction methods will be. We prioritized our application design to maximize the accessibility and minimizing the effort required by the user in capturing audiophotographs.

In our app, we deal with two types of audio recordings: memo and ambient audio. Regarding the timing of when to record the audio, we can consider either before the camera shutter or after the camera shutter, or possibly spanning across it. Our current design records the ambient audio up until the camera shutter, and lets the user record the memo audio on demand. During actual use by a visually impaired participant accompanied by a sighted companion, quite often the time before the camera shutter activation was spent with the sighted companion giving guidance and feedback to the participant on where to position the camera for the best shot. Because such exchanges were being included in the ambient audio recording, a number of participants expressed that the current ambient audio recording may not be of much value as a cue for recalling and describing the context. At the same time, there were also participants who used this functionality to their advantage by letting the device record the sighted companion's voice as he started describing the object or scene that was about to be captured.

Under the photo browsing mode, when perusing through the photos and hearing the associated audio recordings, we designed it such that the memo recording would play first, followed by the ambient audio recording, each time a new

photo was scrolled onto the screen. As mentioned above, when capturing a photo, the moments leading up to the shutter activation is spent adjusting and focusing on the act of capturing the photo, so the ideal time to record a memo description may be after capturing the photo. Given the importance of the memo audio in supporting quick identification of each photo, we follow the principle of presenting necessary information first by playing back the memo audio, followed by the ambient audio. Clearly, there are many other possible combinations of when to record and playback the memo and ambient audio, and further investigation is needed to explore the advantages of each and to design a user interface that allows for maximum flexibility without complicating the required interaction.

The Camera for the Blind design prototype [7] is a stand-alone camera that has a physical button for the shutter, which, when pressed all the way, activates the shutter, and while held down, records the ambient audio. In the application we have developed, we have incorporated custom touch-screen gestures for controlling the camera mode, since relying on standard button-based UI would have required VoiceOver users to have to physically locate the on-screen buttons or to sequentially sift through them, interrupting not only the process of user attempting to capture a photo but also the VoiceOver feedback potentially polluting the audio recording. While custom gestures do require some learning, we found through our study that users were able to quickly learn them and were able to achieve single-gesture access to the key functions.

#### **Audio Recording vs. Text Entry of Photo Description**

Even our participants who mentioned that they were used to assigning text label to photos they have captured on their camera phone found the ability to record audio memo to be much easier and faster and enable more detail to be captured. Of course, there will be situations when entering text label is more desirable rather than recording audio, and we view our memo audio recording approach as an augmentation and not a replacement for such labeling method. Ideally, regardless of whether the user chooses to record via voice or via text entry, the difference should be transparent from the user's perspective. In the case of recording the memo with voice, the user gets the added bonus of contextual information in the form of audio recording of their voice and the background ambience, which they may or may not choose to peruse at a later time, but the underlying description should be treatable as textual data within the application. This will enable consistent handling of such information for various purposes, from being able to perform a keyword search to auto-captioning the photograph with the text description when sharing it via email or online services.

Two participants also specifically expressed the desire to be able to edit or augment the memo recording at a later time. P5 mentioned that as he was showing the photo to the sighted viewer, he gained additional information about the

photos from the viewer that he wished he could incorporate back into the memo of the photos. The ability to modify and augment memos becomes quite challenging when dealing with recorded audio, and especially if the editing needs to be supported in an eyes-free manner. A solution that takes advantage of textual representation of the recorded memo (either automatically recognized or manually annotated by the user) may be desirable. Further research and design iterations are needed to create an ideal solution for eyes-free memo editing and management.

#### **Richer Augmented Photographs**

Given the variety of sensors and network connectivity that are becoming standard on many smartphones, there is much opportunity for capturing richer contextual information automatically, such as detecting whether the user is indoor or outdoor, deducing building names or even what may have been in the view of the camera based on the user's location and compass heading [1], the weather information at the time and place of capture, etc. Automatic image classification and recognition can obviously add many valuable metadata to the photograph. When appropriate, crowdsourcing mechanism as well as friendsourcing [2], which is a method to distribute tasks only to the requester's friends, can play a valuable role in describing photographs.

It would be interesting to see how audio, photographs, video, and audiophotographs will evolve as they become augmented with richer contextual information. As Frohlich speculates [6], each medium will likely continue to serve different needs under various scenarios. Audiophotograph as a medium is still not widely proliferated, and there is a need for tools that support simple and accessible capturing and editing of audiophotographs, as well as standardization of data formats for packaging an audiophotograph as a single unit.

#### **Blind to Blind and Sighted to Blind Sharing**

There has not yet been much investigation into how audiophotographs may be received by people who are visually impaired. At the end of each of our user study sessions, we experimentally showed each participant some audiophotographs that we had prepared to gauge their reaction and hear their comments. As part of the "showing" experience, we played back the audio associated with the audiophotographs as we verbally described to them the context captured. One example of an audiophotograph we showed consisted of a photograph of a festival float with the sound capturing the cacophony of marching band sounds. Most of the participants thought that it would be an interesting way for them to be able to "consume" a photograph, but also expressed that whoever is taking the audiophotograph would need to be quite conscious of capturing "good" sound as well as a good photograph. More investigation will need to be conducted to explore the various issues that need to be considered in supporting sharing among the visually impaired and the sighted.

## CONCLUSION

Photo sharing is a rich activity that enables us to connect and to share our experiences and views of the world around us with one another. As such, it is not an activity that only the sighted are privy to, but can hold significant meaning even to those without the sense of sight. While a visually impaired person may not be able to visually consume what is captured in a photograph, they can capture their own photographs of their context and share them with sighted people. We conducted an online survey to investigate the current photo sharing practices of people with visual impairment, and found that a large majority of the respondents had strong interest in and even experience with taking and attempting to share photographs. We also found that a large number of respondents face a barrier when it comes to organizing and identifying their photos.

To address these issues, we created an accessible smartphone application that enables people with visual impairment to capture audio recordings along with photographs and to peruse them at a later time. Through the use of such audio recordings, a user can associate a spoken memo describing the particular photograph, along with any ambient sound that may aid in recalling the photo later. Our user study with five visually impaired participants in naturalistic settings revealed that our approach is a viable solution to overcome the barrier faced by blind and partially sighted users in sharing their photos. Based on our observations of visually impaired people engaged in the end-to-end process of capturing, browsing, and sharing photographs, we presented a set of insights about future design considerations for making the photo sharing experience more accessible and inclusive.

## ACKNOWLEDGMENTS

We thank our study participants for their time commitment and the valuable feedback, as well as our questionnaire respondents. This work was supported in part by the National Science Foundation under the East Asia and Pacific Summer Institutes for U.S. Graduate Students (EAPSI) award, of which one of the authors (Dustin Adams) was a recipient. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

1. Baber, C., Cross, J., Khaleel, T., and Beale, R. Location-based photography as sense-making. *Proc. BCS-HCI 2008*, British Computer Society (2008), 133–140.
2. Bernstein, M.S., Tan, D., Smith, G., Czerwinski, M., and Horvitz, E. Personalization via crowdsourcing. *ACM Trans. Computer-Human Interaction* 17, 2 (2008), 1–28.
3. Bigham, J.P., Jayant, C., Ji, H., et al. VizWiz: nearly real-time answers to visual questions. *Proc. UIST 2010*, ACM (2010), 333–342.
4. Bigham, J.P., Jayant, C., Miller, A., White, B., and Yeh, T. VizWiz::LocateIt - enabling blind people to locate objects in their environment. *Proc. CVAWI Workshop 2010*, IEEE (2010), 65–72.
5. Frohlich, D. and Tallyn, E. Audiophotography: practice and prospects. *Ext. Abstracts 1999*, ACM (1999), 296–297.
6. Frohlich, D.M. *Audiophotography: Bringing photos to life with sounds*. Kluwer Academic Pub, 2004.
7. Haidary, N. Camera for the Blind. <http://nadeemhaidary.com/camera.html>.
8. Jayant, C., Ji, H., White, S., and Bigham, J.P. Supporting blind photography. *Proc. ASSETS 2011*, ACM (2011), 203–210.
9. Ko, J. and Kim, C. Low cost blur image detection and estimation for mobile devices. *Proc. ICACT 2009*, IEEE Press (2009), 1605–1610.
10. Many Ltd. Picle. <http://www.picleapp.com/>.
11. Robinson, S. Com-Phone Story Maker. <https://play.google.com/store/apps/details?id=ac.robinson.mediaphone>.
12. UCR/California Museum of Photography. Sight Unseen. <http://www.cmp.ucr.edu/exhibitions/sightunseen/>.
13. Vázquez, M. and Steinfeld, A. Helping visually impaired users properly aim a camera. *Proc. ASSETS 2012*, ACM (2012), 95–102.
14. White, S., Ji, H., and Bigham, J.P. EasySnap: real-time audio feedback for blind photography. *Proc. UIST 2010*, ACM (2010), 409–410.
15. Luo Xu. Speature. <http://itunes.apple.com/us/app/speature/id401887412>.
16. Sensory Photography. [http://www.photovoice.org/html/pvmethodology/method\\_04/index.htm](http://www.photovoice.org/html/pvmethodology/method_04/index.htm).
17. Blind with Camera School of Photography. <http://www.blindwithcameraschool.org/>.
18. Apple - Accessibility - VoiceOver - In Depth. <http://www.apple.com/accessibility/voiceover/>.
19. StoryMark Life. <http://www.storymarklife.com/>.
20. Voicepic. <http://voicepic.me/>.
21. Storyrobe. <http://storyrobe.com>.
22. Photos by Blind Photographers - Photo Essays - TIME. [http://www.time.com/time/photogallery/0,29307,1897093\\_1883579,00.html](http://www.time.com/time/photogallery/0,29307,1897093_1883579,00.html).
23. Seeing Beyond Sight: Photographs by Blind Teenagers. Book, video & TED Talk. <http://www.seeingbeyondsight.org/home/>.
24. Blind With Camera. <http://www.blindwithcamera.org/>.
25. Seeing With Photography Collective. <http://www.seeingwithphotography.com/>.
26. oMoby. <http://www.iqengines.com/omoby/>.
27. Seeing with Sound - The vOICe. <http://www.seeingwithsound.com/>.