BSpeak: An Accessible Crowdsourcing Marketplace for Low-Income Blind People

Aditya Vashistha
University of Washington
adityav@cs.washington.edu

Pooja Sethi
University of Washington
psethi17@cs.washington.edu

Richard Anderson
University of Washington
anderson@cs.washington.edu

ABSTRACT
BSpeak is an accessible crowdsourcing marketplace that enables blind people in developing regions to earn money by transcribing audio files through speech. We examine accessibility and usability barriers that 15 first-time users, who are low-income and blind, experienced while completing transcription tasks on *BSpeak* and Mechanical Turk (MTurk). Our mixed-methods analysis revealed severe accessibility barriers in MTurk due to the absence of landmarks, unlabeled UI elements, and improper use of HTML headings. Compared to MTurk, participants found *BSpeak* significantly more accessible and usable, and completed tasks with higher accuracy in lesser time due to its voice-based implementation. In a two-week field deployment of *BSpeak* in India, 24 low-income blind users earned ₹7,310 by completing over 16,000 transcription tasks to yield transcriptions with 87% accuracy. Through our analysis of *BSpeak*'s strengths and weaknesses, we provide recommendations for designing crowdsourcing marketplaces for low-income blind people in resource-constrained settings.

Author Keywords
HCI4D; accessibility; voice; crowdsourcing; India.

INTRODUCTION
According to the World Health Organization, about 90% of the world's 285 million visually impaired people live in low-income settings [2]. India, a fast developing economy, is home to over 63 million visually impaired people—the largest proportionally among all countries [52]. Over 70% of India's visually impaired population live in rural regions [4], almost 50% are illiterate [4], and a majority of them live in poverty [47]. Limited access to educational resources [55] and employment opportunities [51] severely impede their potential to overcome poverty.

Mainstream crowdsourcing marketplaces like Amazon Mechanical Turk (MTurk) [9] and CrowdFlower [13] have enhanced earning potential of people by incentivizing them to perform time-sensitive microtasks, such as image and keyword tagging, translation, and transcription, among others. However, several accessibility and usability barriers, including visual CAPTCHA during sign-up process, incomplete task descriptions, inaccessible task features, inability to find accessible tasks, and time restrictions to complete tasks, limit the adoption of crowdsourcing marketplaces by blind people [60].

In prior work [57], we designed, built, and evaluated *Respeak*—a voice-based, crowd-powered platform to generate transcripts for audio files in local languages and accents spoken in developing regions—and presented results of cognitive experiments and deployment with *sighted* university students in India. A key difference between *Respeak* and other online marketplaces is how *Respeak* users produced transcripts. Instead of typing audio content, they listened to a short audio segment and then repeated the content into an automatic speech recognition (ASR) engine to generate a transcript.

In this work, we draw inspiration from the voice-based design of *Respeak* to create *BSpeak*, an accessible crowdsourcing marketplace for speech transcription to provide additional income-generating opportunities to low-income blind people. In contrast to *Respeak* [57], our focus is to explore how *BSpeak* could enhance the earning potential of low-income *blind* people from rural and peri-urban areas by capitalizing on their speaking and listening skills.

To examine accessibility and usability barriers in *BSpeak* and compare them with those present in a mainstream microtasking platform, we conducted a usability study with 15 blind participants who completed equivalent speech transcription and information retrieval tasks on *BSpeak* and MTurk. Our mixed-methods analysis of task completion rate, task completion time, usability parameters, and semi-structured interviews indicated that blind participants encountered severe accessibility barriers while completing accessible tasks on MTurk. Poorly designed user interface (UI), unlabeled UI elements, improper use of HTML headings, and absence of landmarks exacerbated their user experience. In contrast to MTurk, participants found *BSpeak* significantly more accessible and usable due to its voice-based implementation. They completed significantly more tasks in a lesser time with a higher performance and a lower mental demand, frustration, and effort.

To analyze *BSpeak*'s ability to enhance earning potential, we conducted a two-week field deployment with 24 low-income blind people from rural and peri-urban regions in India. *BSpeak* users collectively completed over 16,000 tasks, repeating 2,560 unique segments with an accuracy of 61.6%

CHI 2018, April 21–26, 2018, Montréal, QC, Canada.

http://dx.doi.org/10.1145/3173574.3173631

to earn ₹7,310 (USD 111)[1]. The expected payout per hour of the application use was ₹36, comparable to the average hourly wage rate in India [16]. To reduce errors in speech transcription, the *BSpeak* engine used multiple string alignment (MSA) and a majority voting process to align transcripts generated by up to eleven users. It produced transcription for Indian English audio files with 87% accuracy, and Hindi files with 92% accuracy. The cost of transcription was USD 1.20 per minute, one-fourth of the industry standard for transcription of audio files in local languages and accents [5–7]. Our mixed-methods analysis indicated that *BSpeak* enhanced earning potential of blind people. Users also self-reported improvements in general knowledge and language skills.

To compare crowdwork performance of blind people to sighted people, we examined *BSpeak*'s use by blind people and *Respeak*'s use by sighted people since *BSpeak* and *Respeak* have the same underlying design. We seeded *BSpeak* with tasks used in *Respeak*'s deployment, and compared the speech transcription performance of blind users and sighted users on different languages and content type. Compared to sighted users, blind users completed three times more tasks and earned 2.5 times more money in just half the time. However, the accuracy of transcription generated by blind users was 11% lower than the accuracy yielded by sighted users due to differences in education level, socioeconomic status, and language skills. The *BSpeak* engine needed to align transcripts generated by eleven blind users to yield accuracy comparable to the alignment of transcripts generated by five *Respeak* users. We discuss the lessons learned from the usability study, field deployment, and comparison study, and offer recommendations for designing accessible crowdsourcing marketplaces for low-income blind people in resource-constrained settings.

RELATED WORK

Visually impaired people face insurmountable challenges in finding employment. For example, over 32% of visually impaired American adults are unemployed [54] and a staggering number are underemployed [53]. Many employers have reported hesitations to hire people with visual impairment due to perceived concerns of higher expenditure and lower productivity [41]. Lack of information, poor self-confidence, limited mobility skills, difficulties in using transportation, and high cost of assistive technologies further reduce employment opportunities for blind people [39]. In developing regions, these employment barriers are compounded by infrastructural constraints, socioeconomic barriers, and low literacy rates. For example, almost 50% of India's blind population is illiterate and about 70% live in rural resource-constrained settings [4]. There is an acute shortage of people with disabilities in the government, public sector, and private sector [1]. Although there is a large body of work examining barriers that impede job prospects of blind people in developed countries [45], only a few scholars have studied employment barriers for blind people in developing regions. The most notable work is by Pal and Lakshmanan who investigated workforce participation of visually impaired assistive technology users in Bangalore (India), and found that several high-skilled visually impaired people were employed in low-paying jobs and faced systematic discrimination [51]. Naraharisetti and Castro conducted a cross-sectional study that revealed gender, disability type, and illiteracy as key factors associated with employment of persons with disabilities in India [49]. Khanna et al. studied the relationship between poverty and blindness in India, and found that blind people from rural and poor families are most at risk [47]. Our work extends this literature by presenting the benefits and limitations of online marketplaces perceived by low-income blind people in India.

There is a large body of research at the intersection of accessibility and crowdsourcing. Bigham et al. designed *VizWiz* that enabled blind people to ask visual questions to a pool of MTurk workers in real time [34]. Brady et al. explored the use of *VizWiz Social* to let blind users ask visual questions from their friends on social media platforms [37], and conducted a longitudinal study to analyze these questions [36]. Since blind users were hesitant to seek help from their social media friends due to the perception of high social costs, Brady et al. designed *Visual Answers*, a Facebook application that posts visual questions by blind people in the News Feed of volunteers [35]. Zhong et al. designed *RegionSpeak* that enabled blind people to receive labels from MTurk workers for all relevant objects contained in a stitched image [59]. Lasecki et al. further improved the *VizWiz* system to design *Chrous:View* that facilitates a real-time conversation between a blind user and crowd workers about a video stream from the user's phone. *Be My Eyes*, a smartphone application with over half a million users in 150 countries, establishes a live video connection to let blind users ask questions from sighted volunteers [11]. It is important to note that blind people are *consumers* of crowdwork in these systems [60]. Since *BSpeak* allows blind people to perform speech transcription tasks, we contribute to this literature by examining the performance of low-income blind people in India as *producers* of crowdwork.

Ours is not the first study to highlight usability and accessibility barriers in MTurk. Khanna et al. examined usability barriers encountered by low-income people with basic digital literacy skills, and suggested using local language, and simplified UI and task instructions [48]. Zyskowski et al. examined challenges faced by MTurk workers with disabilities in the United States at different stages of the crowdwork process [60]. They found that several accessibility barriers, including the inability to create an account because of CAPTCHA, poor ratings due to unfinished inaccessible tasks, and a lack of a filter to select accessible tasks, limited the earning potential of blind MTurk workers. Our work has notable differences from [48] and [60]. In contrast to [48] which examined usability barriers of low-income sighted people and [60] which reported accessibility barriers faced by blind people in the United States, we conducted studies with low-income blind people from rural and peri-urban areas in India. Moreover, while Zyskowski et al. relied on surveys and interviews to identify accessibility barriers [60], we designed a usability study to directly observe the first interactions of blind users with MTurk.

Several crowdsourcing marketplaces have been designed for low-income people in developing regions. Both *mClerk* [46]

[1] In this paper, we use an exchange rate of USD 1 = ₹66 (or INR 66).

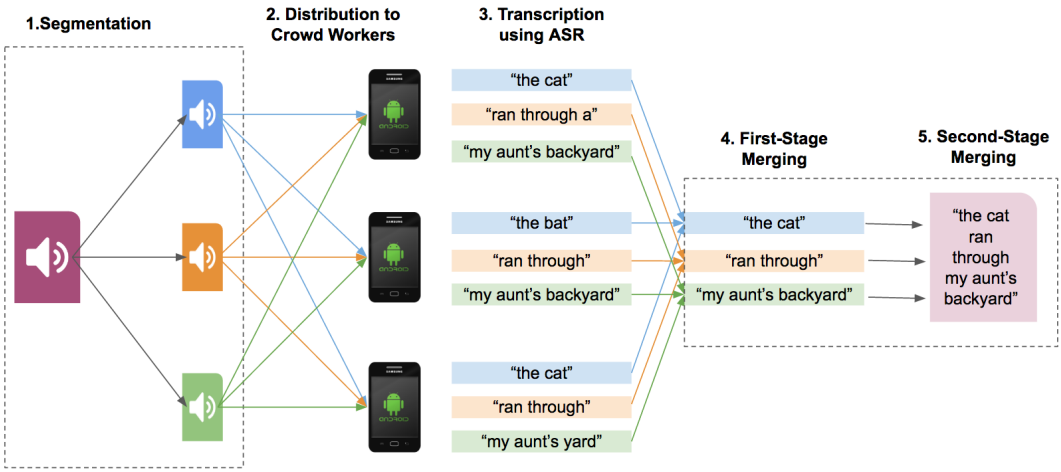

Figure 1: A high-level illustration of *BSpeak*'s design. Areas inside dotted lines represent the processes of the *BSpeak* engine.

and *MobileWorks* [50] allow low-income people to transcribe images sent to their phones. Similarly, *TxtEagle* [42] and *mSurvey* [20] facilitate low-income people to answer SMS-based surveys. *mCent* provides airtime to its users to watch videos and use new smartphone applications [17]. *Samasource* hires low-income people to perform categorization, data mining, and transcription tasks [23]. These crowdsourcing marketplaces, except *mCent*, require workers to have decent reading and typing skills, often in local languages. To provide additional earning opportunities to people with limited typing skills, we created *Respeak*—a voice-based, crowd-powered system to transcribe audio files in local languages and accents—and deployed it with sighted users who produced transcripts by speaking audio content into the ASR-enabled Android application instead of typing it [57]. In this work, we design, build, and evaluate *BSpeak*, an accessible online marketplace that extends *Respeak* and capitalizes on its voice-based implementation. In contrast to *Respeak*, we deploy *BSpeak* to low-income blind people with rural background for evaluating its potential to supplement their income. Ours is also the first work to compare the performance of crowdwork by blind people to sighted people. We now describe the design of *BSpeak* and outline the changes we made to *Respeak* to enhance its accessibility for blind people.

BSPEAK SYSTEM AND APPLICATION DESIGN

Similar to *Respeak*, *BSpeak* uses a five-step process to obtain a transcription for an audio file, as illustrated in Figure 1.

1. **Audio Segmentation:** Based on the speaking rate and occurrence of natural pauses in an audio file, the *BSpeak* engine segments the audio file to yield short segments, typically three to six seconds in length, that are easier for crowd workers to remember.

2. **Distribution to Crowd Workers:** Each audio segment is randomly distributed to multiple *BSpeak* application users.

3. **Transcription by Crowd Workers using ASR:** A *BSpeak* user listens to a segment and then re-speaks the content into the application. The application uses the built-in Android ASR engine to convert the audio input into a transcript that is read aloud to the user. If the transcript is similar to the audio segment, a user submits it in order to receive a new segment. The transcript is expected to have several errors since a user may not fully understand the audio content or the ASR engine could incorrectly recognize some words.

4. **First-Stage Merging:** Once a predefined number of users have submitted their individual transcripts for a particular segment, the *BSpeak* engine merges their transcripts using MSA and a majority voting process to obtain a best estimation transcript. If speech recognition errors are randomly distributed, aligning transcripts reduces the errors since the correct word is recognized for the majority of the users. The *BSpeak* engine then compares individual transcripts to the best estimation transcript to determine users' reward for that particular task. Once the cumulative amount earned by a user reaches ₹10, a digital payment using *Paytm* [21] is sent to the user.

5. **Second-Stage Merging**: The *BSpeak* engine concatenates the best estimation transcript for all segments into one large file to yield the final transcript for the original audio file.

BSpeak App: An Accessible Version of the Respeak App

Blind users could navigate *BSpeak* by using TalkBack—Android's built-in screen reader software—that reads aloud screen content on touch and swipe gestures. As illustrated in Figure 2, we changed several components of *Respeak*'s UI by using Android's accessibility guidelines [8] to create *BSpeak*:

- **Labeling UI elements:** We labeled all UI elements, such as buttons and images, with appropriate descriptions to enable TalkBack to read them aloud. Without such labels, TalkBack

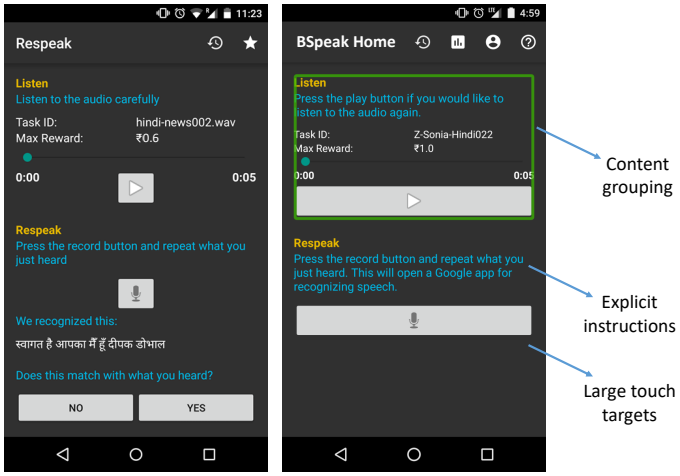

Figure 2: A screenshot of the *Respeak* application home (left) and the *BSpeak* application home (right).

would make generic announcements such as *"image button"* or *"text area"* which provides no information about a UI element's actual functionality.

- **Large touch targets:** Since it is difficult for blind users to navigate small touch targets on a phone's screen, we minimized empty space, increased button and font sizes, and made the touchable areas more than 48dp x 48dp.
- **Explicit instructions:** We modified verbal instructions that complement visual cues accessible only to sighted users. For example, the instructions for recording audio on the *Respeak* home screen were *"Press the record button and repeat what you just heard."* While sighted users could see a separate Google app opening up for speech recognition, blind users needed to be explicitly told to expect it. Thus, in *BSpeak*, we changed the verbal instruction to *"Press the record button and repeat what you just heard. This will open a Google app for recognizing speech."*
- **Content grouping:** We grouped multiple UI elements into single announcements to treat them as one focusable container. Thus, as the user presses any single element within the group, the entire content of the container is announced out loud by TalkBack, which makes it easy to access logically related elements all-at-once. Without grouping, a blind user would have to individually touch text labels or swipe many times to read the elements on the screen.

We also added new features to *BSpeak* such as access to help page and the option to select the language of audio tasks. Another key difference between *Respeak* and *BSpeak* is how users accessed the output transcript generated by the ASR system. While *Respeak* users could read the transcript displayed on the screen, TalkBack read it aloud for *BSpeak* users.

USABILITY EVALUATION OF BSPEAK AND MTURK

We conducted a usability study with low-income blind people to investigate usability and accessibility barriers in *BSpeak*, and compare them with those present in MTurk. The barriers observed during this phase provided key insights to refine *BSpeak*'s design.

Recruiting Participants

To recruit low-income blind participants, we contacted Enable India—a non-profit organization in Karnataka that conducts technical training for blind people and helps them seek employment. The nine-month training program focuses on a wide range of skills development, including English communication, mobility training, lifesaving skills, employability, and basic computer training. The technical component comprises of learning screen reader software, the Microsoft office suite, social media platforms, and other Internet websites. Each training cohort has roughly 35 students. A majority of the trainees come from rural and peri-urban areas, and have poor English skills. Several trainers are alumni of the training program.

Methodology

To ensure that our sample had people with varying technical expertise, we used snowball sampling to recruit 15 low-income blind trainees, trainers, and alumni to participate in the usability study. We randomized and balanced the order in which participants used *BSpeak* and MTurk. For *BSpeak*, we installed the application on an Android phone with the Talkback software. For MTurk, we configured a Windows laptop with JAWS and NVDA screen reader software. Since prior work has reported accessibility barriers for blind people to sign-up and search for accessible tasks on MTurk [60], we created an account and opened the task page for participants before they began a task. For each system, we first provided a brief description and then requested participants to spend ten minutes exploring its user interface. While we did not offer any demonstration of the system upfront, we did provide verbal assistance when participants requested it. Once participants familiarized themselves with the interface, we requested them to complete four tasks. The first two tasks required participants to transcribe a randomly selected short Indian English audio segment, and the other two tasks required them to find the reward amount for the current task and the total amount they had earned.

The MTurk task page had a reward amount, a total amount earned, a collapsible block containing task instructions, and an external link to a webpage that randomly loaded a short audio segment and provided a text area to type the transcript. On the other hand, the *BSpeak* task page contained task instructions, a reward amount, a randomly selected short audio file, and a record button to re-speak the audio content. Participants had to navigate to another page to determine the total amount earned in *BSpeak*. In both systems, the audio files were roughly 1-7 seconds in length, and the next transcription task automatically loaded after the participants submitted the current task.

We recorded the completion status and time taken for each task. Once participants completed all tasks for a system, we conducted interviews to let them score the system on accessibility and NASA TLX parameters, which included mental demand, performance, effort, and frustration. We also asked them to describe the strengths and weaknesses of the system.

Participants then repeated the activities for the other system. Each session lasted approximately 1.5 hours.

Analysis

To determine statistically significant differences between *BSpeak* and MTurk, we used a Wilcoxon signed-rank test or paired samples t-test (based on whether the distribution is parametric) for task completion rate, task completion time, accessibility score, mental demand, performance, effort, and frustration. We recorded and transcribed interviews in English, and subjected our data to thematic analysis as outlined in [38].

Participants Demographics

Nine participants were totally blind and six were partially blind. Participants were 28 years old, on average. Eleven participants were male and four were female. Ten participants had completed a bachelor's degree, four had finished high school, and one participant had been home schooled. Participants came from a variety of vocations: six participants were trainees, five were trainers, and one each was an accessibility tester, operational analyst, solution specialist, and technical support engineer. The average monthly income of employed participants was USD 306. Participants were from seven states in India. Over half had families living in rural and peri-urban regions, and one-third had family members working as farmers and laborers. On average, the annual family income for a family size of 4.2 was USD 4,500. All participants were non-native English speakers, and several reported facing difficulties with English language.

With respect to technology use, all participants had an Android phone and a Windows laptop. They used a wide variety of screen reader software, including NVDA (15), TalkBack (13), JAWS (9), SuperNova (3), Windows Eye (2), and eSpeak (2), among others. None of the participants had prior experience with MTurk, *BSpeak*, and *Respeak*. Five participants tried earning money online earlier on *Paid to Click* websites [3, 14, 24], albeit with no success, due to accessibility issues.

Usability Evaluation Findings

While all participants successfully completed the *BSpeak* tasks, seven participants could not complete one or more tasks on MTurk. A Wilcoxon signed-rank test indicated a significant difference in task completion rate for *BSpeak* and MTurk (W=0, Z=-2.63, p=.015). Participants also took significantly lesser time to complete tasks on *BSpeak* (M=324s, SD=117s) than on MTurk (M=858s, SD=260s), t(14)=9.43, p<.001. Compared to MTurk, participants rated their performance higher on *BSpeak* and experienced lower mental demand, effort, and frustration. A Wilcoxon signed-rank test indicated significant difference between *BSpeak* and MTurk on mental demand (W=105, Z=3.39, p<.001), performance (W=0, Z=-3.42, p<.001), effort (W=88, Z=2.94, p=.001), and frustration (W=104, Z=3.28, p<.001).

Thirteen participants found *BSpeak* more accessible than MTurk, while two found MTurk equally or more accessible than *BSpeak*. The average accessibility score of *BSpeak* and MTurk was 9.1 (SD=1.2) and 6.2 (SD=1.9), respectively. A Wilcoxon signed-rank test indicated significant difference in the scores of *BSpeak* and MTurk (W=1.5, Z=-3.25, p<.001).

Our qualitative analysis indicated that participants struggled with several accessibility barriers in using MTurk. Web Content Accessibility guidelines of W3C [26] recommend using HTML headings to reflect page organization, and emphasize not skipping heading ranks (*e.g.,* <h1> should not be followed by <h3>). Seven participants complained about the improper use of HTML headings in MTurk; the first heading started with <h6>. Though the guidelines also recommend labeling regions [18] or using ARIA landmarks [25] to distinguish page regions, such as main, navigation, search, banner, and application, four participants struggled due to the absence of regions and landmarks. Though MTurk visually organizes data in a tabular format, three participants reported the absence of structural markup that differentiates header and data cells. Four participants could not use access keys, and three participants struggled with unlabeled buttons, text labels, and links. Five participants could not hear the beginning of an audio task due to its overlap with the output of screen reader software. One participant complained about the improper use of link element [15], the use of white text on a blue background, and the inability to search for accessible tasks. A participant, an accessibility tester by profession, stated:

> *UI elements are not labeled. Navigational and content regions are also not specified. Headings are not structured. <h1>, <h2>, <h3> should be there, but it starts with <h6>! Blind people will get confused and say, 'what is this?' I don't think blind people will be able to do work on this. Somehow, I did, but I am not satisfied.*

Participants also encountered significant usability barriers in using MTurk. Ten participants found the user interface complicated, and six participants complained about overwhelming information on the task page. Three participants struggled in accessing the external link due to the sheer number of links on the task page that confused them about which link to click. Other three participants thought they had already clicked on the link since the screen reader software read aloud the URL listed next to the linked content. Five participants accidentally closed the task instructions, implemented as a collapsible block, and read them only when we intervened. Three participants found it challenging to go back and forth between playing the audio file and typing the transcript. Despite our instructions, seven participants did not click 'accept HIT' before beginning the task, and three struggled with MTurk terminologies. A participant stated:

> *I am not able to understand this world [MTurk ecosystem]. What is 'completion code'? They should use simple words that anyone with basic English can understand. The page repeatedly says 'link, link, link.'*

Twelve participants found *BSpeak* accessible and usable. They commended its layout and appreciated that a help feature was available. Eleven participants reported that the ability to transcribe audio files by speaking was *BSpeak*'s key strength. Since some participants were already using voice input to open applications, search information, and call friends on their phone, it was easy for them to transcribe content by speaking into the ASR system. They also preferred speaking to avoid spelling errors that typing might generate due to constrained

Content type	Audio files	Tasks	Duration
Interview	3	275	13.1
Lecture	1	9	0.4
News	2	27	1.4
Phone call	3	37	1.8
Public speech	8	1102	88.7
Song	3	77	7.2
TED talk	3	931	46.5
TV ad	1	10	0.7
YouTube program	3	89	5.4

Table 1: Number of audio files, tasks, and total duration (in minutes) for each content type.

space on the phone's keyboard. Nine participants appreciated the ability to use *BSpeak* anytime and anywhere due to the ubiquity of mobile phones. Four participants considered *BSpeak* a *"learning while earning"* tool that could also improve their pronunciation skills and help them understand new accents—a skill many participants deeply valued. Three participants reported accessibility glitches, such as a misplaced focus on the homepage, an inability to increase font size using pinch-zoom gesture, and the use of black text on a red background in one of the BSpeak page. Five participants were concerned that remembering large audio files and the lower accuracy of ASR could make *BSpeak* difficult to use for some blind people.

Similar to the findings of Pal and Lakshmanan [51], several participants echoed the concerns that blind people face difficulties in finding employment, and that many experience workplace discrimination, limited growth, and low salary. They appreciated the prospect of MTurk and *BSpeak* as an alternative source of income. Several participants valued the ability to *"work from home"* at their *"own pace"* due to mobility and English language barriers. A few of them appreciated that these systems could conceal their disability status, and thus, protect them from potential discrimination. On the other hand, three participants worried that exclusively using these systems could isolate them and result in missed opportunities due to lack of social interactions. Such concerns with the use of MTurk have also been reported by Zyskowski et al. [60].

In summary, enthusiastic responses from participants, and strong quantitative and qualitative results prompted us to conduct a field deployment of *BSpeak* to evaluate its strengths, weaknesses, and potential to provide additional earning opportunities to low-income blind people. We now describe the details of the field deployment in India.

FIELD DEPLOYMENT IN INDIA

Through the network of blind trainees, staff, and alumni of Enable India, we used snowball sampling to recruit 24 low-income blind Android users to participate in a two-week deployment of *BSpeak*. Among these, ten users also participated in the usability study. During a face-to-face orientation session, we installed the application on users' phones, gave them a brief demonstration of *BSpeak*, and collected demographic information. To set the right expectations, we informed all participants that *BSpeak* is a research prototype and our goal is to investigate its accessibility and feasibility, and that there are no immediate plans to release the application widely. At the end of the deployment, we conducted a web survey to evaluate their user experience. The survey had 13 subjective questions that spanned several themes including benefits and limitations of the *BSpeak* application, usability and accessibility barriers, and potential of *BSpeak* to supplement their income.

BSpeak Users Demographics

Eighteen users were totally blind and six were partially blind. Users were 27 years old, on average. Nineteen users were male and five were female. Seventeen users had completed a bachelor's degree, four had completed high school, two had finished middle school, and one had earned a master's degree. Employed users (N=13) had an average monthly income of USD 313. Unemployed users (N=11) were dependent on their family with average monthly income of USD 169 for a family of size four, and thus were living below the poverty line of USD 1.90 per day [22]. Even though all users resided in Bangalore at the time of the study, most of them moved to Bangalore in the last three years from rural and peri-urban regions of nine Indian states. Half of them had family members in rural regions who were either dependent on them (N=3), or were working as farmers and laborers (N=9).

All users had access to mobile Internet, and used TalkBack as the primary screen reader on their phone. Their average monthly phone expense was ₹300 (USD 4.5). None of the users had prior experience with speech transcription and crowdsourcing marketplaces. Users were native speakers of Kannada (N=13), Hindi (N=4), Telugu (N=3), Marathi (N=1), Konkani (N=1), Malayalam (N=1), and Assamese (N=1). Several of them had poor Hindi and English language skills. While the self-reported scores, on a ten-point scale, for local language listening and speaking skills averaged to 9.8, the average scores for English listening, English speaking, Hindi listening, and Hindi speaking skills were 7.2, 7, 6, and 5.9, respectively. During the orientation session, many users made frequent grammatical errors while speaking English and Hindi.

Speech Transcription Tasks

We selected 27 audio files in Indian English and Hindi for transcription by *BSpeak* users. Out of these, 21 audio files were the same as those used in *Respeak*'s deployment because we aimed to compare the accuracy of crowdwork by blind users to sighted users. Since a majority of the deployment participants expressed discomfort with Hindi, we selected more transcription tasks in English than in Hindi. Although the participants were more comfortable in other local languages like Kannada and Telugu, we could not provide tasks in these languages since they are not yet supported by Android's built-in ASR engine. The combined duration of the audio files was 2.75 hours, and it comprised of a wide variety of content including interviews, lectures, news, phone calls, public speeches, songs, TED talks, TV advertisements, and YouTube programs. The *BSpeak* engine segmented these files to yield 2,560 segments (*i.e.,* micro transcription tasks) that varied from three to six seconds in length. We outline the number of audio files, tasks, and total duration for different content types in Table 1 and languages in Table 2.

Language	Audio files	Tasks	Duration
English	14	2060	123 mins
Hindi	13	500	43 mins

Table 2: Number of audio files, tasks, and total duration for each language.

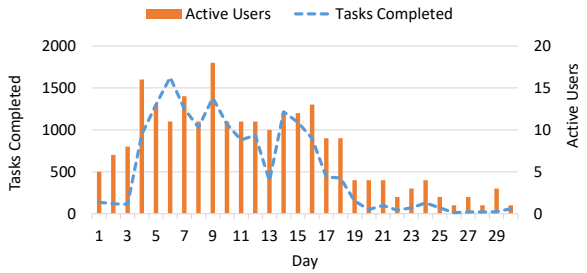

Figure 3: Time series analysis of active users and tasks completed.

Reward Scale and Payment

To ensure that the earning potential of *BSpeak* users equaled that of *Respeak* users, we employed the reward structure used in *Respeak*'s deployment. The maximum reward amount for each task was calculated as ₹0.2 per second. If a user generated a transcript with an accuracy over 80%, the user received the entire reward amount for that task. If the accuracy was below 80% yet over 50%, the user received a maximum reward proportionate to their accuracy. The user received no rewards if the accuracy was below 50%. Since we could not predict how soon and with what accuracy users would complete tasks, we could not estimate the availability and accuracy of the best estimation transcripts. We thus used a pre-computed ground truth to evaluate users' transcripts every ten minutes. The maximum amount a user could earn by completing all Hindi and English tasks was ₹514 (USD 7.8) and ₹1,470 (USD 22.3), respectively. Users could access the reward structure, the amount earned, accuracy, a list of all completed tasks, and the date of last transfer on the *BSpeak* application.

Analysis

We used a mixed-methods approach spanning quantitative analysis of word error rate (WER), cost, and performance, as well as a qualitative analysis of surveys to evaluate *BSpeak*. All but one user completed the survey at the end of the deployment. We subjected their responses to thematic analysis as outlined in [38]. We also compared *BSpeak*'s use by blind people to *Respeak*'s use by sighted people on tasks completed, WER, performance, and transcription cost.

Deployment Findings

BSpeak users used the application for over 208 hours and completed 16,000 tasks, repeating 2,560 segments with an average accuracy of 61.6% to earn ₹7,310 (USD 110). Twelve users completed Hindi tasks 3,133 times, and 24 users completed English tasks 12,872 times. On average, users listened to a segment three times and repeated the content into the ASR system 1.7 times. The median task completion time was 49 seconds. The expected payout per hour of the application use was ₹36, which is comparable to the average hourly wage rate in India [16]. The *BSpeak* engine combined the transcripts generated by eleven users to yield a transcription with 87.1% accuracy and USD 1.20 per minute of transcription cost.

Ground truth:	is	not	as	rosy
Transcript 1:	*its*	not	*at*	rosy
Transcript 2:	is	not	as	*rose*
Transcript 3:	is	—	as	rosy
Majority voting:	is	not	as	rosy

Table 3: Alignment of transcripts obtained from three speakers. Missing words are marked as — and incorrect are *italicized*.

Figure 3 shows the time series analysis of tasks completed and unique active users. Though the deployment ended after two weeks, some users continued using the application for a month. *BSpeak* became very popular among the social network of our users; we received over 20 requests from blind friends of users to access the application. The popularity of *BSpeak* prompted Enable India's management to request a disclaimer from us:

> *We understand that the BSpeak app has become a super hit among our trainees and staff. Most people we have asked love it and also have been making a lot of money through the tasks. We request you to send us a statement that you chose people on your own for deployment based on your criteria, and they are making money through the app based on their skills, and that the organization has shown no favoritism in this.*

Speech Transcription Accuracy and Cost

The average WER for transcripts generated by individual *BSpeak* users was 38.4%. We ran a series of experiments to reduce the WER by aligning transcripts generated by multiple users. For each segment, we conducted ten runs of the experiment. In each run of the experiment, we randomly selected transcripts generated by K users and aligned them using MSA and a majority voting process. We averaged the WER of the best estimation transcript generated in each run. Table 3 shows an example of how WER of an English segment was reduced by aligning transcripts generated by three users.

We used the experimental setup to align transcripts generated by three, five, seven, nine, and eleven users by varying the value of K. Table 4 shows the number of tasks done by K or more users, WER obtained, and transcription cost for alignment of K users. To compute the cost, we multiplied transcription rate (*i.e.,* ₹0.2 per second per user) with expected payout (*i.e.,* 61.6% of the transcription rate). Our analysis indicated that the overall WER decreased as the value of K increased. The cost linearly varied based on the number of transcripts used in the alignment process.

Table 4 also shows the WER obtained for English and Hindi tasks after aligning transcripts generated by K users. The average WER for English tasks was higher than Hindi tasks because all users were non-native English speakers and had no choice but to complete tasks in English. In contrast, only those users who were confident in Hindi opted to complete Hindi tasks. In our survey, several users reported struggling

K	Tasks done by ≥ K users	WER (%)			Cost (USD/min)
		Overall	English	Hindi	
1	2,560	38.4	40.7	30.1	0.1
3	2,509	30.7	33.0	22.6	0.3
5	1,904	26.8	29.9	19.0	0.6
7	833	19.9	20.9	12.1	0.8
9	708	17.6	17.9	11.3	1.0
11	89	12.9	13.0	8.3	1.2

Table 4: WERs and transcription cost obtained after aligning transcripts generated by K users.

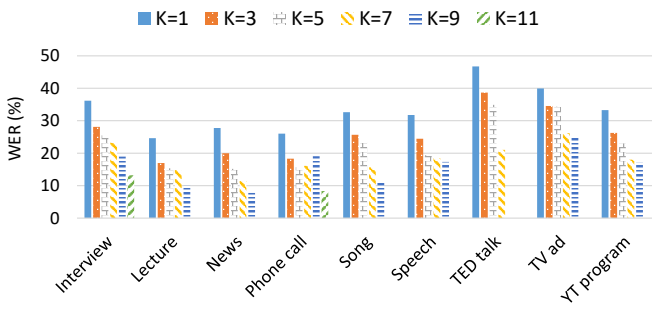

Figure 4: WER obtained after alignment of transcripts generated by K users for different content types. A missing bar indicate that less than K speakers completed tasks.

with English tasks because of *"unfamiliar vocabulary," "fast pace,"* or *"poor recognition by ASR engine."*

Figure 4 plots the WER obtained after aligning the transcripts generated by K users for different content types. The WER for content created for mass consumption like news, lectures, songs, and YouTube programs were low. The WER for public speeches and telephone calls was high because users struggled with ambient noise. The WER for interviews was high due to the prevalence of heavy accents and the overlapping of speakers. Although TED talks and TV advertisements are also created for public consumption, several users struggled with unfamiliar accents and technical terminologies, and yielded poor performance compared to other content types. The analysis of survey responses also indicated that users found speeches (N=11), songs (N=6), and YouTube programs (N=3) easy to remember and re-speak, and faced difficulties in transcribing interviews (N=4), telephone calls (N=4), and TED talks (N=4).

To examine the benefit of requesting blind users to re-speak content into the ASR engine instead of inputting audio files directly to it, we submitted the segments to the Google Cloud Speech API [12]. The average WER of the transcripts thus obtained was 52%—35% more than the average WER obtained by *BSpeak* users. Since blind users repeated audio content into a high-quality microphone of their device in a quiet environment, *BSpeak*'s WER was lower than the WER yielded by the ASR engine on original audio segments.

Financial and Instrumental Benefits

BSpeak users collectively earned ₹7,310. The average amount earned by users was ₹304 (USD 4.60) and the maximum amount earned was ₹1,050 (USD 16). Nine users earned more than their monthly phone expense in just two weeks. Table 5 shows the distribution of the amount earned by *BSpeak* users. All users agreed that *BSpeak* has a strong potential to supplement their earnings. They valued it as a tool to *"utilize their free time in earning money."* Several users earned money for the first time in their life. One such user stated:

Amount earned (₹) ≤	*BSpeak* users
300	16
600	2
900	4
1200	2

Table 5: Distribution of the amount earned by *BSpeak* users.

> *I am grateful to you for creating the app. I earned money for the first time and learned the value of each rupee.*

While 19 users found the amount earned commensurate to the time spent in completing tasks, five users suggested increasing the reward amount to at least ₹1 per task. One of them asserted:

> *Sometimes I feel the reward amount is just okay. You have to increase the amount for each task. ₹0.60 is not acceptable for a low-income person. We also have to consume Internet to use the application.*

While some participants perceived Internet costs to be significant, the total download size of *BSpeak* tasks was 189 MB and the cost was under ₹50—less than 2.5% of the total amount *BSpeak* users could earn by doing tasks. Several users self-reported receiving instrumental benefits such as improved listening skills (N=13), pronunciation (N=11), and concentration (N=3). They found *BSpeak* a *"tool for speech therapy"* and appreciated its ability to introduce them to new accents. Two users reported improvements in their vocabulary by finding the meaning of unfamiliar words they encountered. Another two users indicated that *BSpeak* had increased their knowledge about *"current affairs and new subjects."* Some users stumbled on topics they never explored earlier such as technology and rural inventions, and prison reform movements. Three users appreciated that they could listen to useful content in their free time while earning money. One of them stated:

> *The app improves listening skills, concentration, and pronunciation of difficult words. I listened to old interviews of great personalities like Kiran Bedi, Modi, and Kalam, all while I was earning during my leisure time.*

User Feedback and Preferences

Sixteen users preferred to use the *BSpeak* application at home, four during their commute, and two during office breaks. Most users were active during the evening and at night. Similarly, the usage was much higher on weekends. We also asked users to report constraints that impeded their application use. Five users struggled with the availability of the Internet, five others had unexpected responsibilities at work, two had to travel, and one had to attend a family wedding.

All users commended the accessibility features of *BSpeak*, and found it easy to use and navigate. Eighteen users reported the ability to transcribe files through speech as *BSpeaks*'s key strength since this feature was easy to use, saved time, and

yielded transcripts without spelling errors. Seven users liked that the tasks were not timed and that they could listen to audio files more than once. Five users appreciated the ability to use the application whenever they had time. Five other users found *BSpeak* entertaining to use. One of them exclaimed:

> *I felt like I was playing a game! The freedom to listen to the tasks multiple times helped me to understand fast speech and unclear words. The instant payment also motivated me to complete more tasks.*

Some users faced challenges in using *BSpeak*. Six users found it difficult to get their speech recognized on multiple occasions, especially with homophones. For example, a user stated that even after five attempts, 'phase' was transcribed as 'face'. Eight users found it difficult to remember long sentences that began with a clipped word or contained unfamiliar words.

Users had several recommendations to improve *BSpeak*. Eight users suggested providing a function to edit a transcript through typing after a predefined number of unsuccessful speaking trials. Users also suggested providing a detailed review of the mistakes they had made during previous tasks. They recommended introducing new playback features such as rewind, forward, and manipulate playback speed to improve segment retention. They proposed to include the ability to select tasks based on content type. Some audio segments were blank, had excess noise, or had people clapping. Users were unsure what to re-speak in such scenarios. For example, users recorded *"clapping," "claps," "applause," "noise,"* and *"nothing"* for a segment containing applause. Users recommended adding a feature to report such task anomalies.

Comparison of Blind and Sighted Users

BSpeak is an accessible version of *Respeak* with the same underlying system components and ASR engine, but differences in the design of smartphone application. Since a subset of *BSpeak* tasks were the ones used in *Respeak*'s deployment, we compared the performance of blind users on the tasks that were completed by sighted users in *Respeak*'s deployment.

Table 6 shows the comparison of *BSpeak*'s and *Respeak*'s deployments based on several parameters. Although the duration of *BSpeak*'s deployment was half that of *Respeak*'s deployment, blind users completed over three times more tasks, spent over five times more time on the app, and earned 2.5 times more money than the *Respeak* users. Though blind users were more enthusiastic in doing crowdwork, sighted users generated transcripts with a lower WER than blind users. As a result, the expected payout per hour for blind users was almost half that of the payout for sighted users.

To compare the improvements yielded from MSA and a majority voting process, we used the same experimental setup to align transcripts generated by K users for segments common to both deployments. Table 7 shows that for all values of K, the transcripts generated by sighted users in both Hindi and Indian English yielded a lower WER than the blind users. The differences in WERs of sighted and blind users drastically reduced when transcripts by more users were aligned. The WERs obtained after alignment of transcripts generated by 11 *BSpeak* users were comparable to the WERs obtained after alignment of transcripts generated by five and seven *Respeak* users for Hindi and Indian English, respectively. We also analyzed performance of blind and sighted users on different content types. Both sighted and blind users performed poorly in transcribing interviews and TV advertisements. However, blind users had higher success in transcribing news, lectures, and phone calls. In contrast, *Respeak* users found songs and YouTube programs easiest. In both deployments, the performance of users on speeches was average.

	Respeak	***BSpeak***
Duration	1 month	15 days
Total users	25	24
Unique tasks	756	2,560
Tasks done	5,464	16,005
WER	23.7%	31.2%
Amount earned	₹3,036	₹7,310
Time spent	40 hours	207 hours

Table 6: Comparison of *Respeak* and *BSpeak* on different deployment parameters.

K	***BSpeak* WER (%)**			***Respeak* WER (%)**		
	Overall	**English**	**Hindi**	**Overall**	**English**	**Hindi**
1	31.2	33.7	30.1	21.9	26.2	18.1
3	23.7	26.3	22.6	12.5	18.1	10
5	20.4	23.7	19	10.6	15.2	8.6
7	18.7	21	12.1	10.3	12.3	7.4
9	16.4	17.1	11.3	9.9	11.8	7
11	12.9	13	8.3	9.6	11.4	6.9

Table 7: Comparison of *BSpeak*'s and *Respeak*'s WERs obtained after alignment of transcripts generated by K users.

Disparity in education, socioeconomic status, and language skills contributed to differences in the WER yielded by blind and sighted users. The analysis of demographic information of blind and sighted users revealed that the average education level of sighted users (15.7 years) was higher than the blind users (14.2 years). A Mann-Whitney U test indicated a significant difference in education level of sighted and blind users ($U=173.5$, $Z=-3.023$, $p=.002$). While only one blind user had a master's degree, six sighted users had a master's degree and one was pursuing a Ph.D. Moreover, 20 sighted users had a professional degree compared to only five blind users. Almost half of the blind users were from rural backgrounds while all sighted users were from peri-urban or urban regions. The average monthly family income of blind users was only one-fourth of that of the sighted users. Moreover, while sighted users rated their English and Hindi speaking skills highly, a majority of blind users reported struggling with both English and Hindi, and rated their language skills at best as average.

DISCUSSION AND CONCLUSION

Several researchers have highlighted the constraints blind people face in finding full- and part-time employment [39–41]. While crowdsourcing marketplaces like MTurk have provided additional earning opportunities to over half a million people across the globe [10], severe accessibility barriers impede the use of such platforms by blind people [60]. While our work on *Respeak* highlighted the design of a voice-based microtasking platform which was refined through cognitive exercises and

validated through a field deployment with sighted users [57], in this work, we made three contributions. First, we designed and built *BSpeak*—an accessible voice-based marketplace for blind people—by extending our work on *Respeak*, and evaluated its accessibility through a usability evaluation and feasibility through a field deployment. Through this, we enriched the scholarship on using crowdsourcing to provide additional earning benefits to low-income people and blind people in developing regions [23, 42, 46, 50, 57]. Second, we conducted an in-depth examination of accessibility and usability barriers present in MTurk, and contributed to a growing literature examining accessibility barriers in mainstream social computing systems [43, 56, 58, 60]. Third, to the best of our knowledge, we conducted the first comparison of crowdwork performance of blind people and sighted people.

Our usability study with 15 low-income blind people from rural and peri-urban India indicated that *BSpeak* is significantly more accessible and usable than MTurk. The deployment of *BSpeak* with 24 low-income blind people revealed that *BSpeak* enhanced earning potential of its users who used it enthusiastically to earn ₹7,310 in two-weeks, and produced transcription of Hindi and Indian English audio files with 87% accuracy and at one-fourth of the industry cost. Blind people commended *BSpeak*'s voice-based implementation, and found it easier and faster to produce transcripts without any spelling errors. Users also appreciated the flexibility to complete tasks anytime-anywhere due to *BSpeak*'s mobile-based implementation, and short and untimed tasks. Although the usage of the *BSpeak* application by blind users was much higher than the usage of the *Respeak* application by sighted users, the accuracy of speech transcription by blind users was lesser due to their lower language skills, education, and socioeconomic status than the sighted users.

The errors in transcriptions often resulted from users' mishearing of audio segments and ASR's incorrect recognition of re-spoken content. Another reason that contributed to lower-accuracy transcripts generated by blind users is the fast playback speed of TalkBack. During *BSpeak*'s orientation sessions, we observed that some blind users did not catch minor mistakes in the transcripts generated by the ASR engine since they were read aloud too fast by Talkback. Even when we suggested some users to reduce the playback speed to accurately review generated transcripts, they were hesitant to do so out of the fear of slowing themselves down while using other applications. During the initial phase of the deployment, most errors resulted from users' mishearing of segments and incorrect recognition by the ASR engine. However, as the users gained experience, the ASR errors reduced, the errors due to mishearing of transcripts increased, and the errors due to mishearing of original segments varied based on the clarity and accent of speech in the segments.

The expected payout per hour of *BSpeak*'s use was a modest ₹36 per hour, which is comparable to the average hourly wage rate in India [16], but lower than the minimum wage of skilled workers in urban regions [32]. Although *BSpeak* supplemented the income of blind people, the payout by itself is inadequate to sustain a living in a metropolitan city like Bangalore. Our immediate next step is to explore ways to increase the payout, such as by improving the accuracy with which blind users complete tasks, by decreasing the time taken by them to complete tasks, and by raising the rewards offered for completing tasks. To improve the accuracy, future work could incorporate a functionality to edit a transcript after multiple unsuccessful speaking attempts. Since all but four users were non-native speakers of English and Hindi, future iterations could include tasks in other local languages, such as Kannada and Marathi, by integrating local language speech recognition APIs like [19] for improved performance. To minimize errors due to mishearing of transcripts, the application could either automatically reduce the playback speed of Talkback or send repeated reminders to users to reduce the speed. To decrease the time taken by users to complete tasks, future iterations could automatically skip tasks after a pre-defined number of unsuccessful speaking trials. Sending such difficult tasks to expert users could improve accuracy and reduce task completion time. Since the average industry transcription cost for audio files in local languages and accents is USD 5 per minute [5–7], there is a scope to increase the rate of transcription up to ₹0.8 per second per user while keeping the transcription cost below the industry standard. Though this calculation does not account for server, maintenance, and personnel costs, it could quadruple the payout up to ₹144 per hour. In addition, experimenting with ASR word lattices to reduce WER [44], and sending a segment to more than a predefined number of users only when the transcripts generated by users vary over a threshold could lead to significant reductions in transcription cost, which could result in spillover benefits to users.

Since all participants were literate and employed participants earned more than the national average, more evaluations are needed to validate whether our findings translate to low-literate, illiterate, and underpaid blind people. A promising next step is to scale *BSpeak* by partnering with entities that create high volumes of audio data, such as government helplines [30, 31, 33], and large-scale Interactive Voice Response platforms like Babajob [27], GramVaani [29] and CGNet Swara [28]. More work is needed to design and evaluate advanced task distribution algorithms that maintain data security and confidentiality of audio files containing sensitive information, such as insurance calls and health helplines.

We recommend designers of mainstream crowdsourcing marketplaces to incorporate comprehensible terminologies, design user interfaces within accessibility guidelines, and add a filter to select accessible tasks. We recommend task requesters to provide simple and clear instructions, design accessible content on external webpages, and indicate upfront if a task is inaccessible for blind people. *BSpeak* demonstrated that a simple user interface, use of voice input, and untimed tasks could make a crowdsourcing marketplace more accessible for low-income blind people in resource-constrained settings.

ACKNOWLEDGMENTS

We are grateful to Enable India for helping us recruit visually impaired participants. We thank Moses Chowdary, Rashmi Kanthi, Deepti Natarajan, Jennifer Webster, and the anonymous reviewers for their helpful suggestions.

REFERENCES

1. 2009. *Employment of Disabled People in India*. Technical Report. Diversity and Equal Opportunity Centre (DEOC).
2. 2014. WHO | Visual impairment and blindness. (2014). http://www.who.int/mediacentre/factsheets/fs282/en/.
3. 2016. Get Paid to Click: 10 Legit Sites That Pay You to Click on Ads | SurveySatrap. (Feb. 2016). http://surveysatrap.com/get-paid-to-click/.
4. 2016. *National Institute for the Visually Handicapped Annual Report 2015-16*. Technical Report. Department OF Empowerment of Persons with Disabilities, Ministry of Social Justic and Empowerment, Government of India. http://nivh.gov.in/index.php/annual-report
5. 2016. Quick Transcription Service. (2016). http://www.quicktranscriptionservice.com/Hindi-Transcription.html.
6. 2016. Scripts Complete. (2016). http://scriptscomplete.com/Hindi-Transcription-Services.php.
7. 2016. Transcription Services Us. (2016). http://www.transcription-services-us.com/Language-Transcription-Rates.php.
8. 2017. Accessibility | Android Developers. (2017). https://developer.android.com/guide/topics/ui/accessibility/index.html.
9. 2017. Amazon Mechanical Turk. (2017). https://www.mturk.com/mturk/welcome.
10. 2017. AWS Developer Forums: MTurk CENSUS: About how many workers were on Mechanical Turk in 2010? (2017). https://forums.aws.amazon.com/thread.jspa?threadID=58891.
11. 2017. Be My Eyes ~ Lend Your Eyes to the Blind. (2017). http://bemyeyes.com.
12. 2017. CLOUD SPEECH API: Speech to text conversion powered by machine learning. (2017). https://cloud.google.com/speech/.
13. 2017. CrowdFlower: AI for your business. (2017). https://www.crowdflower.com/.
14. 2017a. Four Dollar Click | Earn Money by Viewing Ads. (2017). https://fourdollarclick.com/.
15. 2017. H59: Using the link element and navigation tools | Techniques for WCAG 2.0. (2017). https://www.w3.org/TR/WCAG20-TECHS/H59.html
16. 2017. India Average Daily Wage Rate Forecast 2016-2020. (2017). http://www.tradingeconomics.com/india/wages/forecast.
17. 2017. Jana. (2017). https://www.jana.com/.
18. 2017. Labeling Regions Page Structure WAI Web Accessibility Tutorials. (2017). https://www.w3.org/WAI/tutorials/page-structure/labels/.
19. 2017. Liv.ai. (2017). https://liv.ai/.
20. 2017. mSurvey. (2017). https://www.msurvey.co/.
21. 2017. Paytm. (2017). https://paytm.com/.
22. 2017b. Poverty & Equity Data | India | The World Bank. (2017). http://povertydata.worldbank.org/poverty/country/IND, urldate = 2016-09-30.
23. 2017. Samasource. (2017). http://www.samasource.org/.
24. 2017c. Two Dollar Click : Get Paid To Click. (2017). http://www.twodollarclick.com/index.php?view=click.
25. 2017. Using ARIA landmarks to identify regions of a page - WCAG WG. (2017). https://www.w3.org/WAI/GL/wiki/Using_ARIA_landmarks_to_identify_regions_of_a_page
26. 2017. WCAG Overview | Web Accessibility Initiative (WAI) | W3C. (2017). https://www.w3.org/WAI/intro/wcag.
27. 2018. Babajob. (2018). http://www.babajob.com/.
28. 2018. CGNet Swara. (2018). http://cgnetswara.org/.
29. 2018. Gram Vaani. (2018). http://www.gramvaani.org/.
30. 2018a. Hotlines & Helplines | National AIDS Control Organization | MoHFW | GoI. (2018). http://naco.gov.in/hotlines-helplines.
31. 2018b. Kisan Call center. (2018). http://dackkms.gov.in.
32. 2018c. Minimum Wages | Chief Labour Commissioner. (2018). http://clc.gov.in/clc/min-wages.
33. 2018d. NCW::Helpline. (2018). http://ncw.nic.in/frmhelpline.aspx.
34. Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. 2010. VizWiz: Nearly Real-time Answers to Visual Questions. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)*. ACM, New York, NY, USA, 333–342. DOI: http://dx.doi.org/10.1145/1866029.1866080
35. Erin Brady, Meredith Ringel Morris, and Jeffrey P. Bigham. 2015. Gauging Receptiveness to Social Microvolunteering. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1055–1064. DOI: http://dx.doi.org/10.1145/2702123.2702329

36. Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013a. Visual Challenges in the Everyday Lives of Blind People. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2117–2126. DOI: http://dx.doi.org/10.1145/2470654.2481291

37. Erin L. Brady, Yu Zhong, Meredith Ringel Morris, and Jeffrey P. Bigham. 2013b. Investigating the Appropriateness of Social Network Question Asking As a Resource for Blind Users. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1225–1236. DOI:http://dx.doi.org/10.1145/2441776.2441915

38. V. Braun and V. Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa

39. Stacy Butler, Adele Crudden, and William Sansing. 2005. Overcoming barriers to employment: Strategies of rehabilitation providers. *Journal of Visual Impairment & Blindness (JVIB)* 99, 06 (2005).

40. Anthony Candela and Karen Wolffe. 2002. A qualitative analysis of employers' experiences with visually impaired workers. *Journal of Visual Impairment & Blindness (JVIB)* 96, 09 (2002).

41. Adele Crudden, Lynn W McBroom, Amy L Skinner, and J Elton Moore. 1998. Comprehensive Examination of Barriers to Employment among Persons Who Are Blind or Visually Impaired. (1998).

42. Nathan Eagle. 2009. Txteagle: Mobile Crowdsourcing. In *Proceedings of HCI International*. Springer-Verlag.

43. Kristin Skeide Fuglerud, Ingvar Tjøstheim, Birkir Rúnar Gunnarsson, and Morten Tollefsen. 2012. Use of Social Media by People with Visual Impairments: Usage Levels, Attitudes and Barriers. In *Proceedings of the 13th International Conference on Computers Helping People with Special Needs - Volume Part I (ICCHP'12)*. Springer-Verlag, Berlin, Heidelberg, 565–572. DOI: http://dx.doi.org/10.1007/978-3-642-31522-0_85

44. Yashesh Gaur, Florian Metze, and Jeffrey P. Bigham. 2016. Manipulating Word Lattices to Incorporate Human Corrections. In *INTERSPEECH*.

45. Yvonne HH Goertz, Brigitte AG van Lierop, Inge Houkes, and Frans JN Nijhuis. 2010. Factors related to the employment of visually impaired persons: A systematic literature review. *Journal of Visual Impairment & Blindness* 104, 7 (2010), 404.

46. Aakar Gupta, William Thies, Edward Cutrell, and Ravin Balakrishnan. 2012. mClerk: Enabling Mobile Crowdsourcing in Developing Regions. In *Proceedings of CHI*.

47. Rohit Khanna, Usha Raman, and Gullapalli N. Rao. 2007. Blindness and poverty in India: the way forward. *Clinical & Experimental Optometry* 90, 6 (Nov. 2007), 406–414. DOI:http://dx.doi.org/10.1111/j.1444-0938.2007.00199.x

48. Shashank Khanna, Aishwarya Ratan, James Davis, and William Thies. 2010. Evaluating and Improving the Usability of Mechanical Turk for Low-income Workers in India. In *Proceedings of the First ACM Symposium on Computing for Development (ACM DEV '10)*. ACM, New York, NY, USA, 12:1–12:10. DOI: http://dx.doi.org/10.1145/1926180.1926195

49. Ramya Naraharisetti and Marcia C. Castro. 2016. Factors associated with persons with disability employment in India: a cross-sectional study. *BMC Public Health* 16 (Oct. 2016), 1063. DOI: http://dx.doi.org/10.1186/s12889-016-3713-6

50. Prayag Narula, David Rolnitzky, and Bjoern Hartmann. 2011. MobileWorks: A Mobile Crowdsourcing Platform for Workers at the Bottom of the Pyramid. In *In Proceedings of HCOMP*.

51. Joyojeet Pal and Meera Lakshmanan. 2012. Assistive Technology and the Employment of People with Vision Impairments in India. In *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development (ICTD '12)*. ACM, New York, NY, USA, 307–317. DOI: http://dx.doi.org/10.1145/2160673.2160711

52. D. Pascolini and S. P. Mariotti. 2012. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology* 96, 5 (May 2012), 614–618. DOI:http://dx.doi.org/10.1136/bjophthalmol-2011-300539

53. PD Rumrill and CM Scheff. 1997. Impact of the ADA on the employment and promotion of persons who are visually impaired. *Occupational Health and Industrial Medicine* 6, 37 (1997), 311.

54. Alexander Shaw, Deborah Gold, and Karen Wolffe. 2007. Employment-Related Experiences of Youths Who Are Visually Impaired: How Are These Youths Faring? *Journal of Visual Impairment & Blindness* v101 (Jan. 2007), 7–21.

55. Aditya Vashistha, Erin Brady, William Thies, and Edward Cutrell. 2014. Educational Content Creation and Sharing by Low-Income Visually Impaired People in India. In *Proceedings of the Fifth ACM Symposium on Computing for Development (ACM DEV-5 '14)*. ACM, New York, NY, USA, 63–72. DOI: http://dx.doi.org/10.1145/2674377.2674385

56. Aditya Vashistha, Edward Cutrell, Nicola Dell, and Richard Anderson. 2015. Social Media Platforms for Low-Income Blind People in India. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. ACM, New York, NY, USA, 259–272. DOI: http://dx.doi.org/10.1145/2700648.2809858

57. Aditya Vashistha, Pooja Sethi, and Richard Anderson. 2017. Respeak: A Voice-based, Crowd-powered Speech Transcription System. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1855–1866. DOI: http://dx.doi.org/10.1145/3025453.3025640

58. Brian Wentz and Jonathan Lazar. 2011. Are Separate Interfaces Inherently Unequal?: An Evaluation with Blind Users of the Usability of Two Interfaces for a Social Networking Platform. In *Proceedings of the 2011 iConference (iConference '11)*. ACM, New York, NY, USA, 91–97. DOI: http://dx.doi.org/10.1145/1940761.1940774

59. Yu Zhong, Walter S. Lasecki, Erin Brady, and Jeffrey P. Bigham. 2015. RegionSpeak: Quick Comprehensive Spatial Descriptions of Complex Images for Blind Users. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2353–2362. DOI: http://dx.doi.org/10.1145/2702123.2702437

60. Kathryn Zyskowski, Meredith Ringel Morris, Jeffrey P. Bigham, Mary L. Gray, and Shaun K. Kane. 2015. Accessible Crowdwork?: Understanding the Value in and Challenge of Microtask Employment for People with Disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 1682–1693. DOI: http://dx.doi.org/10.1145/2675133.2675158