Designing Voice Interfaces for Accessible Crowdwork

Alex C Williams
University of Waterloo
Waterloo, ON N2L3G1, CA
alex.williams@uwaterloo.ca

Edward Lank
Univeristy of Waterloo
Waterloo, ON N2L3G1, CA
lank@uwaterloo.ca

Alexandra Vtyurina
University of Waterloo
Waterloo, ON N2L3G1, CA
avtyurin@uwaterloo.ca

Edith Law
Univeristy of Waterloo
Waterloo, ON N2L3G1, CA
edith.law@uwaterloo.ca

CSCW'18, November 3–7, 2018, Jersey City, New Jersey.

DOI string from ACM form confirmation

Abstract
Today, voice-based systems and interfaces are prevalent in households across the world, yet these systems lack the utility to function as working interfaces for a range of common and fundamental tasks in crowdsourcing marketplaces. In this paper, we describe the current challenges associated with using voice interfaces for crowdwork from the collective lens of crowdworkers with visual impairments. To that end, we propose a brief research agenda at the intersection of accessibility, voice interfaces, and crowdwork. We conclude by highlighting the need for collaborations between scholars in computer-supported cooperative work, human-computer interaction, and other areas of research to reach practical solutions to the presented challenges.

Author Keywords
Crowdsourcing, voice interfaces, accessibility.

ACM Classification Keywords
H.5.m [Info. interfaces and presentation (e.g., HCI)]: Misc.

Introduction
Crowdwork is a relatively new paradigm of online work that leverages the services or abilities of individuals in addressing problems unsolvable by machines. Over the course of the past decade, a number of companies have launched

platforms (e.g., Amazon Mechanical Turk[1], CrowdFlower[2], and UpWork[3]) that instantiate this work paradigm, creating and maintaining a global, geographically distributed workforce for arbitrary work [7]. Work on these platforms range from simple, brief tasks (e.g. audio transcription) to complex, long-term commitments (e.g. collaborative design work). Recent research has shown that these platforms' populations have continued to grow with time, suggesting that crowdwork is a labor paradigm that is here to stay [4].

Despite the work paradigm's prevalence, the vast majority of crowdwork relies on human senses (i.e. sight and hearing), which makes crowdwork largely inaccessible to individuals with visual or hearing impairments. Further, while a range of tasks (e.g., surveys, annotation, editing) do exist on crowdsourcing platforms, modern web browsers and the tooling to support crowdworkers with disabilities leave much to be desired. However, research has shown that *voice interfaces* may be a promising pathway for supporting individuals with impairments when other forms of technology fail to provide accessible experiences [8]. Collectively, prior research poses a critical research question at the intersection of voice interfaces, accessibility, and crowdwork:

How can voice interfaces facilitate crowdwork for individuals with disabilities and impairments?

In this paper, we argue that the current state of accessibility for crowdwork should be recognized as a grand challenge for the design of voice-based systems and interfaces.

[1] https://www.mturk.com
[2] https://www.crowdflower.com
[3] https://www.upwork.com

Voice Interfaces for Accessible Crowdwork

Accessibility is a thriving area of research in human-computer interaction research. However, the body of research focusing specifically on *crowdworkers with disabilities* is relatively small. The literature that does exist suggests that these individuals often use voice interfaces in unique and unexpected ways [1, 2, 8]. Here, we provide context for a concrete research thrust within the domain of *accessible crowdwork*. While our discussion is limited to the following themes, our overarching research agenda is relevant to the future of accessibility in crowdwork at large.

Vision Impairment

We focus our discussion on one common type of disability: *vision impairment*. Vision impairment is among the most commonly reported disabilities, affecting more than 940 million people in the world [9]. We have chosen vision impairments for our discussion as we believe it highlights the most challenging use-case for voice interfaces in crowdwork.

Audio Transcription

We further ground our discussion in the context of audio transcription for two reasons. First, the accessibility community consistently recognizes "audio-only" transcription as an important area of study, specifically for studying older adults [3]. Second, research has shown that individuals may have heightened sensory inputs (i.e., hearing) in the presence of a sensory impairment (i.e., vision). [2, 6].

Voice Assistant Speakers

We situate our suggestions for future work in the context of *voice assistant speakers* (e.g., Amazon Alexa, Google Home, and Microsoft Cortana Invoke). We focus specifically on dimensions of voice interfaces in the context of *performing crowdwork*. It should be noted that other important considerations exist at the platform-level (e.g. accepting a HIT via a voice interface) that are not discussed here.

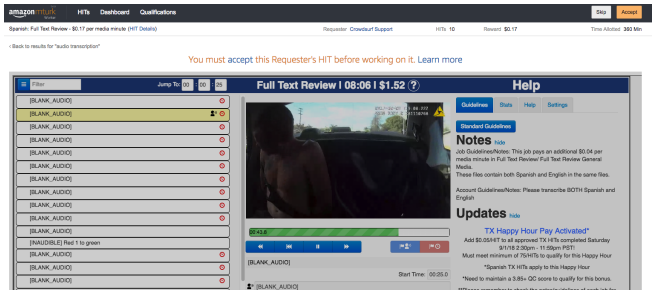

Figure 1: An audio transcription HIT with an accompanied video.

In this context, we can now discuss a brief research agenda with a refined research question:

How can a voice assistant speaker support individuals with disabilities in performing an audio transcription task?

A Research Agenda for Accessible Crowdwork

Given the context described in the prior section, we discuss 3 important directions for the future of accessibility in crowdwork relevant to designing voice assistant speakers: (1) task management, (2) question-answering, and (3) teamwork. These directions were chosen to not only enable visually-impaired crowdworkers, but to *empower* their ability to complete crowdsourced tasks. While important first steps have been taken by Mozilla Research's "Voice-Enabled Web" initiative[4], an important consideration for each of the following points is that the tooling to support voice interactions for web navigation is not natively supported in most modern web browsers.

[4] http://www.it4nextgen.com/mozilla-scout/

I. Designing for Voice-Enabled Task Navigation
First, we propose that future work should focus on providing visually-impaired crowdworkers with the ability to control and navigate task interfaces effectively. Crowdsourcing task interfaces are centered around the submission of new data generated by the user, and include only one or two interface elements (see Figure 1). Beyond screen readers, voice interfaces should be capable of controlling the task interface in a fashion similar to that of standard input devices, i.e. mouse and keyboard, and utilizing different *web-page representations* (e.g., designed specifically for voice interfaces) in support of the task at hand. An important consideration for this particular direction of research is that it relies heavily on manipulating the Document Object Model (DOM) of task webpages, a process mediated by an individuals' browser.

II. Designing for Voice-Enabled Question-Answering
Second, we propose that future work should focus on understanding how individuals with impairments inquire about the task they are performing. Figure 1 shows that the audio transcription task is accompanied by multiple pieces of metadata: a video, task instructions, and information on earning bonus payments. Voice-based interfaces should be capable of answering questions both within and outside of the scope of the task being performed. For example, in the context of Figure 1, a voice interface should be capable of telling the user about the current frame of the video (e.g., the video's timestamp, the objects in the video frame).

III. Designing for Voice-Enabled Teamwork
We propose that future work should explore teamwork between voice interfaces and visually-impaired crowdworkers. This research should be inspired by principles of mixed-initiative interaction [5] in which elements of task delegation, task effort, and subtask sequence are shared between the crowdworker and the voice interface. A simple instan-

tiation of such teamwork, for example, would involve the voice interface recommending task actions in a particular sequence. This direction of research holds promise not only for visual impairments, but also for other disabilities, such as memory impairment [8].

Discussion

In this paper, we discuss important challenges and directions for future research at the intersection of accessibility, voice interfaces, and crowdwork. Like many research initiatives focused on accessibility, the directions discussed in this paper have important benefits for individuals without impairments or disabilities as well, e.g. enabling crowdwork through speakers when other work machines aren't available. Collectively, we recognize the area of accessibility for crowdworkers as a fruitful area of research that deserves more attention and focus from the CSCW community.

While we have described key first steps toward resolving these challenges at a high level, we argue that more robust and long-term solutions require an interdisciplinary community of scholars including computer scientists alongside researchers from the social sciences. During the workshop, we expect to engage with other CSCW scholars and look forward to establishing collaborations for concrete future work.

REFERENCES

1. R. Abdolrahmani et al. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM Press, Galway, Ireland.
2. D. Bragg et al. 2018. A Large Inclusive Study of Human Listening Rates. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM Press, New York, New York, USA, 1–12.
3. R. Brewer et al. 2016. An Approach to Audio-Only Editing for Visually Impaired Seniors. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM Press, New York, New York, USA, 307–308.
4. D. Difallah et al. 2018. Demographics and Dynamics of Mechanical Turk Workers. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM '18*. ACM Press, New York, New York, USA, 135–143.
5. E. Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, New York, New York, USA, 159–166.
6. C. M. Karns et al. 2012. Altered Cross-Modal Processing in the Primary Auditory Cortex of Congenitally Deaf Adults: A Visual-Somatosensory fMRI Study with a Double-Flash Illusion. *Journal of Neuroscience* 32, 28 (7 2012), 9626–9638.
7. A Kittur, J Nickerson, and M Bernstein. 2013. The Future of Crowd Work. *Proc. CSCW '13* (2013), 1–17.
8. Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–13.
9. Theo Vos et al. 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990âĂŞ2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet* 388, 10053 (10 2016), 1545–1602.