Human-Backed Access Technology

Jeffrey P. Bigham, Erin Brady, and Samuel White
Department of Computer Science
University of Rochester
<jbigham,brady,swhite>@cs.rochester.edu

ABSTRACT
In this paper, we introduce the idea of *human-backed access technology*, and outline the motivation for backing up the fragile automatic technology used in access technology with people. We argue that a better understanding of such technology and its interactions with users is needed, and list a number of potential areas in which it would be useful to have explicit guidelines to help direct design and facilitate discussion. Finally, we overview several of our projects in this space, and the research questions that we are trying to answer through them.

INTRODUCTION
The past few decades have seen the development of wonderful new computing technology that serves as sensors onto an inaccessible world for disabled people – as examples, optical character recognition (OCR) makes printed text available to blind people [6, 10], speech recognition makes spoken language available to deaf people [23, 24, 29], and way-finding systems help keep people with cognitive impairments on track [19, 20]. Despite advances, this technology remains both too prone to errors and too limited in the scope of problems it can reliably solve to address the problems faced by disabled people in their everyday lives [15, 16, 29, 30].

Some technology has made it out to people, but still has high error rates in real situations. For example, OCR seems like a solved problem until it fails to decipher the text on a road sign captured by a cell phone camera [21], object recognition works reasonably well until the camera is held by a blind person [5, 15, 25], and the laudable 99% accuracy reported by commercial automatic speech recognition systems [30] falls off precipitously on casual conversation or any time it hasn't been trained for the speaker [29]. Even the automatic techniques used by the screen reading software to convey the contents of the computer screen to blind people are error-prone, unreliable, and, therefore, confusing [7, 8, 17], leading them to be used by only technically-savvy blind people [27]. When access technology is unreliable, it is abandoned [12, 13, 22].

Our work on *human-backed access technology* is about the idea that access technology tools (and intelligent user interfaces in general) would be more reliable and useful if human workers, volunteers, and friends could quickly back up fragile automatic techniques. This approach is directly inspired by how many disabled people already solve problems when their existing strategies fail: ask a friend for assistance [3, 15]. Unfortunately, someone isn't always there to help and asking frequently can make one feel like a burden. Some paid expert services are available, such as captioning services, but they can be expensive and must be arranged in advance. Consequently, they are generally inappropriate for the few seconds of help required to fill in the gaps of existing technology. A blind participant in one of our studies said [15]: *"I get so frustrated when I need sighted help and no one is there."* If access tools were backed up by always-available human-powered services, technology could be more reliable and more useful.

The idea of backing up access technology with humans has been previously articulated [9, 11, 31, 32], but two recent trends have made it practical. First, mainstream mobile phones with low-latency, high-bandwidth connections and a wealth of sensors (camera, microphone, GPS, etc.) have become commonplace, obviating the need for special hardware and making communication on-the-go faster [28]. Second, marketplaces for small jobs like Mechanical Turk [32] and social networks like Facebook [2] and Twitter [1] have grown in popularity [18] providing large pools of potential workers already connected and available in nearly real-time [26]. To fully benefit from this approach, research is needed to better understand how user intelligence, artificial intelligence, and human computation can and should work together into interactive systems.

VizWiz: Our initial work in this area is encapsulated in a system called VizWiz [4]. Blind users of VizWiz are able to take a picture, speak a question, and have it answered by workers on Amazon's Mechanical Turk quickly and cheaply. Our accompanying *quikTurkit* helps improve the response time of Mechanical Turk using several strategies, primarily by pre-queuing multiple workers in advance of receiving the question. We deployed this system to 11 blind people for one week, and used it as a way to understand better what blind people may want from their automatic technology. We also gained insight into how people may want to interact with systems employing interactive human computation. For instance, once receiving an answer, many of our users wanted to ask follow-up questions to the answerer.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

DESIGN IMPLICATIONS

The design of interactive intelligent systems that include humans in the loop has thus far been primarily *ad hoc*. Although best practices are emerging from the point examples that have been developed, we do not yet have explicit design guidelines to assist people working in this space and do not have clearly-defined dimensions on which such systems can be evaluated and compared. The availability of such dimensions in the design of general user interfaces and in the design of intelligent user interfaces [14] has greatly facilitated discussion and driven innovation in these areas. The following are some of the inter-connected areas in which we believe intelligent interactive systems with humans in the loop may differ from what has come before, along with questions we are trying to answer:

- **Privacy and Anonymity**:
  When humans are included in the loop in interactive systems, how can an interface make privacy/anonymity guarantees to users, and how can the tradeoffs in these dimensions be transparently and understandably conveyed to users?
- **Latency:**
  Different sources of human computation may have different expected latencies. How can systems reduce latency? What type of expected latencies are appropriate for different types of work? How can systems help users understand and make decisions based on expected latencies of different sources of human computation?
- **Accuracy:**
  Human computation can provide incorrect answers for a number of reasons, including workers misunderstanding the question, malicious workers, or underspecified questions. How should systems attempt to ensure accurate answers and help convey good estimates of answer quality to users?
- **Feedback to Users:**
  Providing feedback to users about the human computation is that occurring on their behalf is critical for them to make informed decisions. What information do users want or need? How can systems be created to provide this necessary information?
- **Sources of Computation:**
  Users will increasingly face the challenge of deciding between new sources of human computation and artificial intelligence. These sources may differ in terms of cost, availability, and along most of the qualities listed above (e.g. latency, accuracy, privacy, etc). How should systems convey to users the tradeoffs between the different sources of computation currently available to them?
- **Worker Interface:**
  Many of the areas just described are dependent on the work interface. For instance, the design of the worker interface may lead workers to respond more quickly or more slowly. The interface may reveal more private information about the user who submitted the work, or encourage accurate answers. If workers are part of an interactive system, what responsibility do they have for the side effects of their work (e.g. giving a disabled user feedback that causes them harm)? How can the interface convey to the worker the potential side effects of their answers (and potentially their culpability)?

CURRENT PROJECTS

We are exploring these issues within the context of the following interactive systems that we have created that incorporate human computation and artificial intelligence:

VizWiz Social

VizWiz Social extends VizWiz to new sources of human answers, and has the goal of exploring many of the issues outlined in the last section in a real setting. It allows blind users to easily send video questions to Mechanical Turk, Facebook, Twitter, and to contacts over email. We are currently conducting a longitudinal study with VizWiz Social in order to (i) better understand the tradeoffs that people make when choosing sources for answers, (ii) quantify observable differences between these sources (e.g. latency, accuracy), and (iii) investigate different methods of feedback to help users make more informed choices.

VizWiz::LocateIt

VizWiz::Locate it is a project designed to help blind people locate objects in their field of view (e.g. a cereal box on a grocery store shelf, or a particular CD from within a large collection). Users first take an overview picture of the area and speak the name of item they are interested in finding (e.g. "wheaties"). Both the picture and sound file are sent to Mechanical Turk where workers are asked to outline the item in the photograph (a very difficult task to do automatically in general). Computer vision descriptors are generated from the outline and sent to the client device, which uses them to generate audio guidance to help users find their requested item.

This project explores the tight integration of artificial intelligence and human computation, in which human computation is used to solve problems that artificial intelligence cannot yet do.

Monocle

Monocle is an application for the iPhone that lets blind users take a picture and hear text in the photograph read out loud. OCR software running on the iPhone makes a first pass at deciphering the text in the photograph, but either the user or the program may choose to send pieces of the photo off to humans on Mechnical Turk for description. With this project, we are exploring the ability of users to mediate between automatic and human computation.

JEFFREY P. BIGHAM'S BIO

Jeffrey is interested in applying real-time human computation to make intelligent user interfaces more useful and rethinking access technology to make it more available, affordable, and usable for people with disabilities. He received his B.S.E degree in Computer Science from Princeton University in 2003. Starting in fall 2003, he attended the University of Washington, where he worked with Richard E. Ladner as part of the WebInSight research group. He received his Ph.D. in 2009 in Computer Science and Engineering from the University of Washington, and began as an Assistant Professor of Computer Science at the University of Rochester later that year.

CONCLUSION

Intelligent technology can be made more useful in the everyday lives of people if it is backed by human services. Creating effective interactive systems that use both human and artificial intelligence requires a greater understanding of this type of interface along a number of dimensions, some of which we have outlined here. Although we believe this type of interactive system is particularly well-suited for the domain of access technology, most intelligent user interfaces would be more reliable and more useful in our everyday lives if they were backed up by humans.

REFERENCES

1. Amazon's Mechanical Turk. Available from: http://mturk.com.
2. Bernstein, M., Miller, R.C., Little, G., Ackerman, M., Hartmann, B., Karger, D.R., and Panovich, K., Soylent: A Word Processor with a Crowd Inside. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2010)*. 2010.
3. Bigham, J.P., et al., WebinSitu: a comparative analysis of blind and sighted browsing behavior, in Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility. 2007, ACM: Tempe, Arizona, USA. p. 51-58.
4. Bigham, J.P., Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh, Vizwiz: Nearly Real-time Answers to Visual Questions, in Proceedings of the 23rd Annual ACM Symposium on User Interface Software Technology (UIST 2010). 2010.
5. Bigham, J.P., et al., VizWiz::LocateIt - Enabling Blind People to Locate Objects in their Environment, in International Workshop on Computer Vision Applications for the Visually Impaired (CVAVI 2010). 2010, IEEE Press: San Francisco, CA.
6. Bigham, J.P., et al., WebInSight:: making web images accessible, in Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility. 2006, ACM: Portland, Oregon, USA. p. 181-188.
7. Borodin, Y., et al., More than meets the eye: a survey of screen-reader browsing strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*. p. 1-10.
8. Borodin, Y., et al. What's new?: making web page updates accessible. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. 2008. 145-152.
9. Brabyn, J.A., W.F. Crandall, and W.A. Gerrey, Remote Reading System for the Blind: A Potential Application of Virtual Presence. In International Conference on Engineering in Medicine and Biology Society. 1992: Paris, France. 1538-1539.
10. Brudvik, J.T., et al. Hunting for headings: sighted labeling vs. automatic classification of headings. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (ASSETS 2008)*. 201-208.
11. Crandall, W.F., W.A. Gerrey, and J.A. Brabyn, A Fax-Based Reading System for the Blind and Print-Handicapped, in RESNA 19982. 1992: Toronto, Canada. 463-465.
12. Dawe, M., Desperately seeking simplicity: how young adults with cognitive disabilities and their families adopt assistive technologies. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006. 1143-1152.
13. Goette, T., Factors leading to the successful use of voice recognition technology, in Proceedings of the third international ACM conference on Assistive technologies. 1998, ACM: Marina del Rey, California, United States. 189-196.
14. Horvitz, E. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit (CHI 1999)*. 1999, ACM: Pittsburgh, Pennsylvania, United States. 159-166.
15. Jeffrey P. Bigham, C.J., Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samuel White, and Tom Yeh, Vizwiz: Nearly Real-time Answers to Visual Questions, in Proceedings of the 23rd Annual ACM Syposium on User Interface Software Technology (UIST 2010). 2010.
16. Kane, S.K., et al., Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities, in Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility. 2009, ACM: Pittsburgh, Pennsylvania, USA. p. 115-122.
17. Kawanaka, S., et al., Accessibility commons: a metadata infrastructure for web accessibility, in Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility. 2008, ACM: Halifax, Nova Scotia, Canada. p. 153-160.
18. Kittur, A., E.H. Chi, and B. Suh, Crowdsourcing user studies with Mechanical Turk, in Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. 2008, ACM: Florence, Italy. p. 453-456.
19. Liu, A.L., et al., Customizing directions in an automated wayfinding system for individuals with cognitive impairment, in Proceedings of the 11th international

ACM SIGACCESS conference on Computers and accessibility. 2009, ACM: Pittsburgh, Pennsylvania, USA. p. 27-34.

20. Liu, A.L., et al., Indoor wayfinding:: developing a functional interface for individuals with cognitive impairments, in Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility. 2006, ACM: Portland, Oregon, USA. p. 95-102.
21. Manduchi, R., J. Coughlan, and V. Ivanchenko, Search Strategies of Visually Impaired Persons Using a Camera Phone Wayfinding System, in Proceedings of the 11th international conference on Computers Helping People with Special Needs. 2008, Springer-Verlag: linz, Austria. p. 1135-1140.
22. Phillips, B. and H. Zhao, Predictors of Assistive Technology Abandonment. Assistive Technology, 1993. 5.1: p. 36-45.
23. Rabiner, L. and B. Juang, Fundamentals of Speech Recognition. 1993: Prentice Hall PTR.
24. Ravishandkar, M. Efficient algorithms for speech recognition. 2005, Carnegie Mellon University.
25. Samuel White, H.J., and Jeffrey P. Bigham, EasySnap: Enabling Blind People to Take Photographs. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2010)* - Demos. 2010
26. Sorokin, A. and D. Forsyth, Utility data annotation with Amazon Mechanical Turk. Computer Vision and Pattern Recognition Workshops, 2008.
27. Takagi, H., et al. Collaborative web accessibility improvement: challenges and possibilities. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility (ASSETS 2009)*. 2009, 195-202.
28. Union, I.T., Trends in Telecommunication Reform 2007: The Road to NGN. 2007.
29. Wald, M., et al., Correcting automatic speech recognition captioning errors in real time. International Journal of Speech Technology, 2007. 10(1): p. 1-15.
30. Ye, Q., et al., Text detection and restoration in natural scene images. J. Vis. Comun. Image Represent., 2007. 18(6): p. 504-513.
31. Zimmermann, G. and G. Vanderheiden, State of Science: Internet-Based Personal Services on Demand, in Emerging and Accessible Telecommunications, Information and Healthcare Technologies. 2002, RESNA Press. p. 62-70.
32. Zimmermann, G. and G. Vanderheiden, Translation on Demand Anytime and Anywhere, in CSUN's Sixteenth Annual International Conference. 2001: Los Angeles, CA.