Spoken Control of Existing Mobile Interfaces With the Crowd

Walter S. Lasecki
Computer Science Dept.
University of Rochester
Rochester, NY 14627
wlasecki@cs.rochester.edu

Jeffrey P. Bigham
Computer Science Dept.
University of Rochester
Rochester, NY 14627
jbigham@cs.rochester.edu

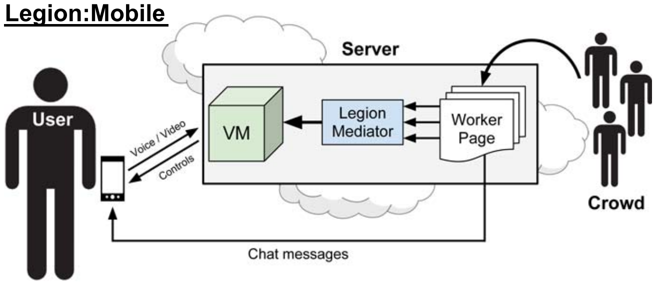


Figure 1: Legion:Mobile is a a conversational assistant that is able to control a user's mobile device when asked to. Using a virtual machine (VM) hosted in the cloud, along with a remote desktop application to provide access, workers can jointly control the device based on verbal requests made by the user. To protect privacy, the user can pause or terminate the session at will to avoid requiring workers to enter sensitive information.

CHI 2013 Mobile Accessibility Workshop, April 28, 2013, Paris, France.

Abstract

Recently, a number of mobile crowd-powered assistive technology systems have been made possible due to the advent of real-time crowdsourcing and ubiquitous broadband internet access on phones and other mobile devices. However, until now these systems have typically aided accessibility *from* the user's device, not *for* it.

In this paper, we design Legion:Mobile, a system that will allow blind and motor impaired users to control their mobile device using more robust spoken controls than are currently available using automatic systems today. We do this by building on prior work on Legion, which allowed the crowd to control existing desktop interfaces, and Chorus, a conversational assistant powered by the crowd. We discuss the design and architecture that makes Legion:Mobile possible, and address privacy concerns often associated with crowd assistance.

Author Keywords

Conversational interaction; real-time crowdsourcing; human computation; intelligent agents

ACM Classification Keywords

H.5.m [Information interfaces and presentation]: Miscellaneous.

General Terms

Human Factors; Design; Economics

Introduction

Over the past few years, a number of mobile crowd-powered assistive technology systems have been developed to helping deaf, blind, and motor-impaired users in their daily lives [2, 5, 6, 4]. These new systems have been made possible by the advent of real-time crowdsourcing and easy access to broadband internet from phones and other mobile devices. However, until now these systems have typically provided accessibility functionality *from* the user's device, not *for* it.

In this paper, we present the design of *Legion:Mobile* , a system that provides more robust spoken controls than are currently available using automatic methods to users who are unable to easily interact with touchscreen- or text-base mobile devices. Legion:Mobile is useful to a number of user groups that are not always considered in mainstream design processes such as blind and low vision, older, and low-literacy users, as well as traditional users who are situationally disabled (i.e. when driving).

We begin by discussing the prior work that Legion:Mobile builds on: (i) Legion, a system that allows the crowd to control existing desktop interfaces, and (ii) Chorus, a conversational assistant powered by the crowd. We then present the architecture of the new system, and discuss some of the privacy issues related to crowd assistance.

Background

Legion:Mobile builds on work in human computation and real-time crowdsourcing. Crowdsourcing is a form of human computation [7] that leverages groups of human workers (often recruited from services such as Mechanical Turk) to solve problems that current automated systems struggle with. Most existing methods derive quality work from redundancy, asking multiple workers to contribute and verify results at each stage. However, these approaches take time, making them ill-suited for interface control. We focus on real-time human computation, which is made possible by recent work showing web workers can be recruited to a task within seconds [1, 2].

Legion: Crowd Control of Existing Interfaces

Legion [5] introduced the idea of continuous real-time crowdsourcing, in order to enable control of existing desktop user interfaces. *Continuous crowdsourcing* recruits workers are asked to complete longer, on-going tasks for as long as they are willing, instead of small individual pieces, allowing workers to maintain context and react to feedback they receive. Legion's *input mediator* merges simultaneous input from multiple workers into a single control stream. This is done by comparing the input of all workers and electing a single leader who is most representative of the decisions the crowd has made in the past. This leader has a very short term length (only around one second) before another vote is taken and a new leader can be selected if need be.

Legion used this model to allows multiple workers to synchronously control an interface as if they were the only a single individual. This unification is key to controlling GUIs without the need for modifications. Legion supports both mouse and keyboard input mediation, but is more effective when discrete actions are presented to workers because it is easier to find consensus. Mobile devices present a particularly well-suited domain to use crowd of control because most are touchscreen based, meaning discrete presses are the most common type of interaction (though behaviors will still present challenges).

Chorus: Conversational Interaction with the Crowd

Chorus [4] is a system that allows users to hold a consistent, reliable conversation with the crowd. Chorus uses an incentive mechanism to encourage workers to collaboratively search for and refine answers, and a shared memory interface to help workers ensure that the current conversation is consistent with previous interactions. Tests have demonstrated that this system can reliably use the crowd to act as a personal assistant and answer user questions via an instant messenger. Legion:Mobile uses a similar interface for chatting, but allows users to speak with the crowd using their voice, which is presented to workers as a playable audio clip in a chat line.

System

Legion:Mobile builds on work on Legion and Chorus, but introduces a new architecture that allows for reliable conversational control of discrete interfaces such as those found on most mobile devices. Users begin by opening the application, which automatically starts recruiting workers from Mechanical Turk (or other crowd). By beginning to recruit workers immediately, the time that the user must wait when the make a request is reduced later.

Conversational Interface

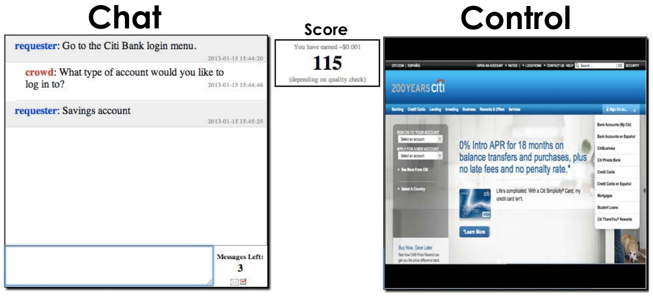
Once the application is running, users are able to use a gesture of their choice on any screen (such as swiping a finger off of a corner, or holding the home button) to start recording a message and then tap to send it. The crowd workers connected to the task will then be presented with a video of the user's screen (similar to remote desktop) and a chat window containing past responses to the user, current response proposals that can be voted on, and recorded user messages that can be played back by clicking on them. The recorded message chat lines also contain a transcript of their content as captioned by automatic speech recognition (ASR). ASR is not always accurate in real-world settings, but can serve to indicate general topic for easy visual search of the chat history.

Crowd Control

In order to allow the crowd to control the user's device remotely, we use mobile remote desktop software available for Android [3]. Our approach can be extended to any such platform with remote desktop software. A virtual machine (VM) instance running on the Legion:Mobile server run the remote desktop client application can be operated via Legion, allowing the crowd to control it. This lets crowd workers connect to a web interface that they interact with using their mouse and keyboard but have access to the user's mobile device.

Worker Interface

Legion:Mobile's worker interface (Figure 2) must be designed to allow workers to switch easily between their two roles. In their first role, workers are asked to hold a conversation with the user to determine what task needs to be completed (this is simple if the user clearly specifies the task and no clarification is needed). In their second role, workers are asked to collectively complete the task on the user's mobile device. This is enabled by providing a remote desktop connection to the user's device on the right side of the screen. Users can interact with this interface as if it were any other application, and their input is forwarded back to the user's device.



requester: Go to the Citi Bank login menu. 2013-01-15 15:44:20

crowd: What type of account would you like to log in to? 2013-01-15 15:44:46

requester: Savings account 2013-01-15 15:45:25

Messages Left: 3

You have earned ~$0.001
115
(depending on quality check)

200YEARS citi
0% Intro APR for 18 months on balance transfers and purchases, plus no late fees and no penalty rate.*
Life's complicated. With a Citi Simplicity® Card, my credit card isn't.
Learn More
Buy Now, Save Later

Figure 2: Layout of Legion:Mobile's two-role worker interface.

Discussion and Future Work

Chorus allows Legion:Mobile to handle spoken controls more robustly than existing systems because it leverages the understanding and contextual understanding of the workers. It can also handle complex or multi-step requests, provided workers understand the sequence of steps needed to complete them. This broad range of capability allows Legion:Mobile to assist users in a way previously not possible using automated system alone, but also introduces privacy concerns.

Privacy

Legion:Mobile shares several of the same privacy limitations that any spoken language interaction or crowdsourcing system does. For example, users must be careful about saying or entering private information when Legion:Mobile is active both because people in their surroundings may overhear them (as with any spoken language interface), and because the crowd workers themselves will then have access to the information. However, since most modern mobile device platforms include privacy protection from people looking over the user's shoulder (an by extension the crowd viewing the screen), this is often less of a concern. For instance, when entering a password, the characters are displayed as dots or stars, and the hover caption for each letter is not shown when typing. Since that is all the crowd would see even in an active session, the risk is largely mitigated.

To be even more secure, the user can divide the task into secure and public portions. For example, users could request that the crowd get them to the login box for a given service, then suspend Legion:Mobile and enter their login information themselves, before resuming the process. This prevents crowd workers from ever having a chance to see private information. Future work will also aim to reduce the amount of information that a single worker can be privy to, even when user's make mistakes in preserving their own privacy (i.e. limiting the maximum time and number of sessions that a worker can assist a user).

Conclusion

In this paper, we have presented Legion:Mobile, a robust conversational control interface for touchscreen mobile devices that leverages real-time crowdsourcing to provide interactive support. We present an architecture for controlling the existing applications and functionality of a user's mobile device with the crowd. Legion:Mobile has the potential to significantly increase the fluidity of interaction not only for blind and motor-impaired users, but also traditional users who are situationally disabled.

References

[1] Bernstein, M. S., Karger, D. R., Miller, R. C., and Brandt, J. R. Analytic methods for optimizing realtime crowdsourcing. In *Proc. of Collective Intelligence*, CI 2012.

[2] Bigham, J. P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R. C., Miller, R., Tatarowicz, A., White, B., White, S., and Yeh, T. Vizwiz: nearly real-time answers to visual questions. In *Proc. of the symp. on User interface software and technology (UIST 2010)*, 333–342.

[3] Damian, K. Remote desktop - google play store. https://play.google.com/store/apps/details?id=pl.androiddev.mobiletab&hl=en. Accessed: 1/08/2013.

[4] Lasecki, W., Kulkarni, A., Wesley, R., Nichols, J., Hu, C., Allen, J., and Bigham, J. Chorus: Letting the crowd speak with one voice. In *In University of Rochester Technical Report* (2012).

[5] Lasecki, W., Murray, K., White, S., Miller, R. C., and Bigham, J. P. Real-time crowd control of existing interfaces. In *Proc. of the symp. on User interface software and technology (UIST 2011)*, 23–32.

[6] Lasecki, W. S., Miller, C. D., Sadilek, A., Abumoussa, A., Borrello, D., Kushalnagar, R., and Bigham, J. P. Real-time captioning by groups of non-experts. In *In Proc. of the symp. on User Interface Software and Technology (UIST 2012)*.

[7] von Ahn, L. *Human Computation*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 2005.