Making Everyday Interfaces Accessible

Tactile Overlays by and for Blind People

Anhong Guo
Carnegie Mellon University

Jeffrey P. Bigham
Carnegie Mellon University

Department editor:
Stephanie Mueller;
stefanie.mueller@mit.edu

Making a physical environment accessible to blind people generally requires sighted assistance. VizLens and Facade put blind users at the center of a crowdsourced, computer-vision-based workflow that lets them make the environment accessible on their own terms.

The world is full of physical interfaces that are inaccessible to blind people. Microwaves, toasters, and coffee machines help us prepare food; printers, fax machines, and copiers help us work; and checkout terminals, public kiosks, and remote controls help us with daily activities. Despite their ubiquity, few interfaces are self-voicing or have tactile labels. As a result, blind people cannot easily use them—they typically must rely on sighted assistance either to use the interface or to label it with tactile markings. However, tactile markings often cannot be added to interfaces on public devices, such as those in an office kitchenette or checkout kiosk at the grocery store, and static labels cannot make dynamic interfaces accessible. Sighted assistance may not always be available, and relying on co-located sighted assistance reduces independence.

Making physical interfaces accessible has been a long-standing challenge.[1,2] Solutions have generally involved making a device self-voicing, modifying the interface (for example, adding tactile markers), or using computer vision (CV). Accessibility can be built into new devices, but not all devices are likely to be covered due to cost. In addition, CV approaches are usually brittle and interface or task specific.[2] The Internet of Things might remedy this situation eventually; as more devices are interconnected and can be controlled remotely, the problem becomes one of digital accessibility, which is easier to solve. For example, users might access a device through a customized interface on their own smartphone.[3,4] However, current solutions are unlikely to make most new physical interfaces accessible or address the significant legacy problem even in the medium term.

We conducted a pair of user studies to better understand how blind people currently use and adapt to everyday home and office appliances. To meet such users' needs, we first built VizLens,[5] a robust and interactive screen reader for real-world interfaces. VizLens users take a

Published by the IEEE Computer Society
1536-1268/18/$33.00 ©2018 IEEE

picture of a desirable interface, which is interpreted quickly by multiple crowd workers in parallel; the system then uses CV to provide real-time feedback and guidance on using the interface. However, VizLens will never be as good as an interface built from the start to be accessible. We therefore created Facade,[6] which combines a crowdsourced interpretation pipeline with CV and a 3D printing application to let blind users produce tactile overlays that turn inaccessible interfaces into accessible ones.

UNDERSTANDING USER NEEDS

To understand blind users' needs, we first went to the home of a blind individual and observed how she cooked a meal and used home appliances. We then conducted semi-structured interviews with six blind participants about their appliance use and strategies for accommodating inaccessible appliances.

Participants felt that interfaces are becoming much less accessible as flat digital touchpads replace physical buttons, which can at least be easily found by fingers once the locations of different functions are memorized. They did not generally have problems locating the control area of the appliances, but they did have problems finding the specific buttons contained within it. We identified four design requirements for a system to generate augmented physical interfaces for nonvisual access.

First, blind people often rely on in-person sighted assistance to identify the original functions and apply the labels for home appliances. The solution for tactile labeling should let blind users independently augment and access their appliances. Second, blind people have different preferences for labeling their appliances, such as reading medium or what buttons to label. The solution should also support rich tactile feedback, diverse labeling strategies, and preferences to address a wide range of individual needs. Third, because blind people cannot remember all the abbreviations and functions, they often choose to only label a few functions, which limits their access. The solution should allow for learning and memorization of the interface. Finally, when the labels they applied wear off because of repeated use, which happens a lot with kitchen appliances, they lose access to the functions and need help again. The solution should support easy attachment and easy reproduction. Of note, static tactile markings cannot make dynamic interfaces accessible. Many interfaces include dynamic components, such as self-checkout terminals in which the full screen changes when the user presses a button. For these types of interfaces, a dynamic solution is required.

VIZLENS: A ROBUST AND INTERACTIVE SCREEN READER FOR REAL-WORLD INTERFACES

Just as the first digital screen readers interpreted the visual information a computer seeks to display,[7] VizLens interprets the visual information of existing physical interfaces (see Figure 1). To work robustly, it combines on-demand crowdsourcing like that found in VizWiz[8] and RegionSpeak[9] with real-time CV such as that used by Access Lens[10] and VizWiz::LocateIt.[11] The system's workflow tightly integrates the strengths of the end user (knowledge of the problem and context, and access to the interface), the crowd (sight and general intelligence), and CV (speed and scalability).

When blind users encounter an inaccessible interface for the first time, they use a smartphone camera to capture a picture of the device and then send it to the crowd. This picture becomes a reference image. Within a few minutes, crowd workers rate the image quality (whether it is blurry and whether the interface area is partially cropped), mark the layout of the interface, annotate its elements (for example, buttons or other controls), and describe each element (for example, labeling buttons "baked potato" or "start/pause"). These results are combined using majority vote. Later, when blind users want to use the interface, they open the VizLens mobile app, point the phone toward the interface, and hover a finger over the screen. CV matches the crowd-labeled reference image to the real-time image. VizLens detects what element the user is pointing at and uses instantaneous audio feedback to guide the user.

Figure 1. VizLens users take a picture of an interface they would like to use, such as that of a microwave oven. This image is interpreted quickly by crowd workers in parallel. The system then uses computer vision to give instantaneous interactive feedback and guidance on using the interface through a mobile (left) or wearable device (right).

A user study of VizLens we conducted with 10 blind participants found the crowdsourced labeling workflow to be fast (8 minutes), accurate (99.7 percent), and inexpensive ($1.15). Based on the participants' feedback, we added functionalities to VizLens that enable it to adapt to interfaces that change state (common with touchscreen interfaces), read dynamic information with crowd-assisted optical character recognition (OCR), and benefit from head-mounted cameras. The resulting system is nearly as robust as a person in interpreting the user interface and nearly as quick and low-cost as a pure CV system, making it feasible to deploy today.

FACADE: AUTO-GENERATING TACTILE INTERFACES FOR APPLIANCES

Two downsides of VizLens are that real-time CV is computationally expensive, and users will not always want to hold or wear the device. We therefore created a way for blind users to independently produce tactile overlays to physical interfaces. Built atop VizLens, Facade adds 3D-printed augmentation to buttons on the original panel (Figure 2).

Figure 2. Facade's crowdsourced fabrication pipeline enables blind people to independently make physical interfaces accessible by producing a 3D-printed overlay of tactile buttons.

Blind users first capture a photo of an inaccessible interface along with a readily available fiducial marker such as a dollar bill as a size reference. The Facade mobile app uses CV to provide aiming feedback to the user in real time. After the user takes a picture, the backend server can warp the image to the front perspective and retrieve the size information. Similar to VizLens, this image is sent to crowd workers to rate its quality and to segment the area that contains the interface. Crowd workers are asked to segment the interface region aligned with the control panel's physical boundaries to make it easier for blind users to attach the overlay by themselves. In parallel, the crowd workers then label the individual buttons by drawing a bounding box of each button and providing a text annotation for it.

Users can then customize the buttons' shape and labels using the Facade mobile app. A virtual version of the interface is displayed on the touchscreen that supports learning and memorization of the interface. Users can also set preferences for reading medium, word abbreviation strategy, whether to include a legend on the side, and so on. Using the crowd-generated labels and users' preferences, Facade then generates a 3D model for a layer of tactile, pressable buttons that fits over the original controls.

Finally, a home 3D printer or mail-order 3D printing service can be used to fabricate the layer, which the blind person aligns and attaches to the interface. Printed overlays from a service such as 3D Hubs (https://3dhubs.com) cost about $10, and they will get cheaper over time. Figure 3 shows several printed overlays and legends generated by Facade with various 3D printing material combinations.

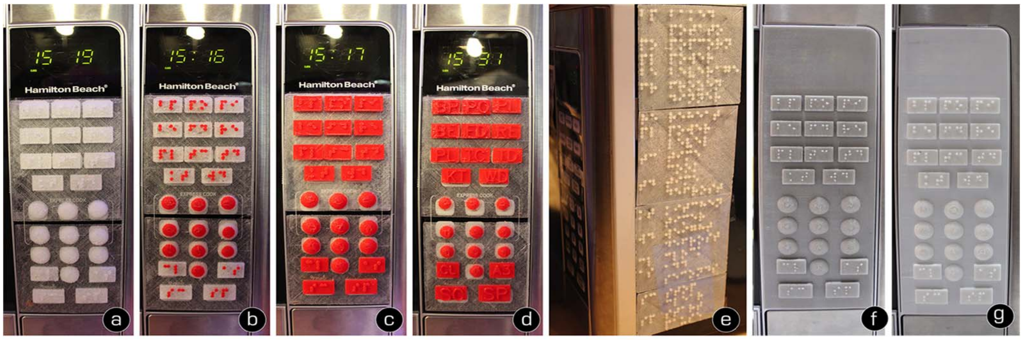

Figure 3. Example printed tactile overlays and legends generated by Facade. (a)–(d) Different 3D printing material combinations tested in system design iterations: NinjaFlex Braille label, NinjaFlex + polylactic acid (PLA) Braille label, NinjaFlex + PLA Braille cover, and NinjaFlex + PLA embossed letter cover. (e) Facade users can choose to print a legend for the abbreviations. (f) and (g) If a user does not have a 3D printer at home, overlays can be printed by a commercial mail-order service such as 3D Hubs using PolyFlex or SemiFlex materials.

A study with 11 blind participants showed that Facade enables blind people to independently create and affix tactile overlays for appliance interfaces, and that the overlays provide rich and usable feedback. Comparing Facade with the traditional method of applying Braille labels, one participant commented that "I like [Facade] much better. I can do it myself, to me it's huge. I don't need to wait for someone to come over and label things for me. If the template gets damaged, I can create a new one. With the [traditional] labels I made, things start peeling off soon. I think this is neat."

Unlike traditional embossed labelers, Facade does not require in-person sighted assistance, provides richer tactile feedback using different reading mediums and button shapes, and reduces memory load by providing a legend and in-app support. In addition, as home 3D printers become faster, blind users will be able to use Facade to generate tactile overlays to home appliances in just minutes.

CONCLUSION

VizLens and Facade solve the long-standing challenge of making everyday interfaces accessible by tightly integrating the complementary strengths of the end user, the crowd, CV, and fabrication technology.

We envision that the Facade approach can work at scale. For example, a sighted partner or a building manager can quickly collect images and automatically produce tactile labels and augmentations to make a space accessible, which is more efficient than manual labeling. Facade also makes end-user fabrication more inclusive by enabling blind users themselves to be makers in the 3D modeling process, shifting the inaccessible component to a crowd interpretation pipeline and an automated model generation tool.

REFERENCES

1. T. Morris et al., “Clearspeech: A Display Reader for the Visually Handicapped,” *IEEE Trans. Neural Systems and Rehabilitation Eng.*, vol. 14, no. 4, 2006, pp. 492–500.
2. G. Fusco et al., “Using Computer Vision to Access Appliance Displays,” *Proc. 16th Int’l ACM SIGACCESS Conf. Computers & Accessibility* (ASSETS 14), 2014, pp. 281–282.
3. G. Vanderheiden and J. Treviranus, “Creating a Global Public Inclusive Infrastructure,” *Proc. 6th Int’l Conf. Universal Access in Human-Computer Interaction* (UAHCI 11), 2011, pp. 517–526.
4. A.A. de Freitas et al., “Snap-To-It: A User-Inspired Platform for Opportunistic Device Interactions,” *Proc. 2016 CHI Conf. Human Factors in Computing Systems* (CHI 16), 2016, pp. 5909–5920.
5. A. Guo et al., “VizLens: A Robust and Interactive Screen Reader for Interfaces in the Real World,” *Proc. 29th Ann. Symp. User Interface Software and Technology* (UIST 16), 2016, pp. 651–664.
6. A. Guo et al., “Facade: Auto-generating Tactile Interfaces to Appliances,” *Proc. 2017 CHI Conf. Human Factors in Computing Systems* (CHI 17), 2017, pp. 5826-5838.
7. J. Thatcher, “Screen Reader/2: Access to OS/2 and the Graphical User Interface,” *Proc. 1st Ann. ACM Conf. Assistive Technologies* (ASSETS 94), 1994, pp. 39–46.
8. J.P. Bigham et al., “VizWiz: Nearly Real-Time Answers to Visual Questions,” *Proc. 23nd Ann. ACM Symp. User Interface Software and Technology* (UIST 10), 2010, pp. 333–342.
9. Y. Zhong et al., “RegionSpeak: Quick Comprehensive Spatial Descriptions of Complex Images for Blind Users,” *Proc. 33rd Ann. ACM Conf. Human Factors in Computing Systems* (CHI 15), 2015, pp. 2353–2362.
10. S.K. Kane, B. Frey, and J.O. Wobbrock, “Access Lens: A Gesture-Based Screen Reader for Real-World Documents,” *Proc. 2013 ACM SIGCHI Conf. Human Factors in Computing Systems* (CHI 13), 2013, pp. 347-350.
11. J.P. Bigham et al., “VizWiz:: LocateIt—Enabling Blind People to Locate Objects in Their Environment,” *Proc. 2010 IEEE Computer Society Computer Vision and Pattern Recognition Workshops* (CVPR 10), 2010, pp. 65-72.

ABOUT THE AUTHORS

Anhong Guo is a PhD Student at Carnegie Mellon University’s Human–Computer Interaction (HCI) Institute. Contact him at anhongg@cs.cmu.edu.

Jeffrey P. Bigham is an associate professor at Carnegie Mellon University’s HCI Institute and Language Technologies Institute. Contact him at jbigham@cs.cmu.edu.