Towards Accessible Audio Labeling of 3D Objects

James M. Coughlan, Huiying Shen, Brandon Biggs
The Smith-Kettlewell Eye Research Institute
2318 Fillmore St.

coughlan@ski.org, hshen@ski.org, brandon.biggs@ski.org

Abstract

We describe a new approach to audio labeling of 3D objects such as appliances, 3D models and maps that enables a visually impaired person to audio label objects. Our approach to audio labeling is called CamIO, a smartphone app that issues audio labels when the user points to a *hotspot* (a location of interest on an object) with a handheld stylus viewed by the smartphone camera. The CamIO app allows a user to create a new hotspot location by pointing at the location with a second stylus and recording a personalized audio label for the hotspot. In contrast with other audio labeling approaches that require the object of interest to be constructed of special materials, 3D printed, or equipped with special sensors, CamIO works with virtually any rigid object and requires only a smartphone, a paper barcode pattern mounted to the object of interest, and two inexpensive styluses. Moreover, our approach allows a visually impaired user to create audio labels independently. We describe a co-design performed with six blind participants exploring how they label objects in their daily lives and a study with the participants demonstrating the feasibility of CamIO for providing accessible audio labeling.

Keywords

Blindness, Low Vision, Visually Impaired, Accessibility, Audio Labeling

Introduction

Many people who are blind or visually impaired have insufficient access to a range of everyday objects, including printed documents, maps, infographics, appliances and 3D models used in STEM education, needed for daily activities in schools, the home and the workplace. While braille labeling is often a useful means of providing access to such objects, there is only limited space for braille, and this method is only accessible to braille readers.

Audio labels are a powerful supplement or alternative to braille labels. They can be created in a variety of ways, including placing touch-sensitive sensors at locations of interest, overlaying a tactile graphic on a touch-sensitive tablet (Talking Tactile Tablet), using a camera-enabled "smart pen" that works with special materials (Miele), or using a stylus such as the PenFriend (Kendrick) that senses small RFID labels affixed to the object. While effective, these audio labeling methods require each object to be customized with special hardware and materials, which is costly and limits their adoption.

By contrast, computer vision-based approaches, in which a camera tracks the user's hands or handheld pointer as they explore an object, enable audio labeling for existing objects with minimal or no customization. Past computer vision approaches to audio labeling include (Shi et al.), which focuses on 3D printed objects, (Thévin and Brock), which uses a depth camera to access flat documents, and (Fusco and Morash), which provides access to tactile graphics to students in educational settings.

We build on our past work on CamIO (Shen et al.; Coughlan and Miele), which we have implemented as a stand-alone iPhone app that tracks an inexpensive wooden stylus held by the user to point at locations of interest on an object. Our approach offers the advantage of enabling the 3D stylus tip location to be estimated with a conventional iPhone camera, which allows us to

develop a novel accessible annotation process. Finally, we report experiments with this annotation method that enables a visually impaired user to define new hotspots and record their own audio labels using a second stylus.

Few of the existing approaches for audio labeling, including earlier versions of CamIO, allow independent labeling for visually impaired users. Although the circumstances when labels are created often require sighted assistance, it is important for a visually impaired user to be in control of the labeling process so they are aware of the labels and their locations. Ensuring visually impaired users have the ability to label also allows them to be on a video call or use an app like Be My Eyes (Weiss) for help creating labels.

Discussion

The next subsections explain how CamIO works and describes the co-design and user study we performed with six blind participants.

Overview of the CamIO Approach

CamIO (short for "Camera Input-Output") is a computer vision-based system for annotating physical objects that provides real-time audio feedback in response to the location on an object that the user is pointing to. The new version (Fig. 1) uses two passive wooden styluses for pointing (see motivation later in this subsection), one for reading hotspots and one for writing them. Another barcode pattern, called the *board*, is mounted to the object, in a location that doesn't occlude the object surface (e.g., underneath the object). The board enables CamIO to estimate the object's position and rotation and thereby determine the stylus tip's 3D location relative to the object.

A hotspot is a 3D point of interest on the object, along with a corresponding text or audio label. In previous versions of CamIO, hotspots were defined in advance by the experimenter, but the new version allows the user to define new hotspots with the writing stylus.

While using a stylus is often less convenient than simply pointing by fingertip, the stylus approach offers several advantages. First, many hand and finger tracking approaches are either unreliable (e.g., relying heavily on skin color cues to segment the hand (Mascetti), and thus easily confused by background objects with a similar color) or require expensive dedicated hardware such as the VR/AR platforms (Oculus Quest); by contrast, computer vision algorithms track an inexpensive stylus reliably over a range of lighting conditions and backgrounds. Second, the 3D location and 3D orientation of the stylus can be estimated with a conventional iPhone camera, which both eliminates depth ambiguities (e.g., the stylus is touching the object vs. hovering several inches above it) and allows us to develop an accessible annotation process. Finally, only part of the stylus need be visible to the camera, which means that the stylus has fewer visibility problems (e.g., the stylus tip location can be estimated even when the tip is invisible because it is contacting a recessed area of the object) than a fingertip tracking approach.

CamIO System Details

The current CamIO system (Fig. 1) runs as a stand-alone iOS app. The object being explored is rigidly attached to the CamIO board, which is a piece of paper with an array of black-and-white barcodes. The iOS device (an iPhone 8 is used in this paper) is mounted on a tripod or held in some fashion to capture the entire object and board. The user holds either a reading stylus or a writing stylus to interact with the object; each stylus consists of a square wooden dowel with a pointy tip, 6 inches long and ½ inch wide, and a paper barcode pattern wrapping all four sides of the dowel. (The reading and writing styluses have different barcodes on them and are thus

distinguishable to the system; we added a rubber band to the top of the writing stylus so that the user can distinguish it from the reading stylus.) Barcode recognition algorithms from the OpenCV software package (OpenCV) determine the locations of all visible barcodes on the stylus and the board, and the Perspective-n-Point algorithm (Szeliski) uses these locations to estimate the 6-dimensional pose (3D location and 3D orientation) of the stylus and board. In this way, the location of the stylus tip is determined in 3D relative to the board and object, which is how CamIO knows where the user is pointing.

Fig. 1. CamIO system overview. A 3D biological plant cell model is mounted on the CamIO board barcode pattern. The reading and writing styluses (the writing stylus has a red rubber band on its end) are next to it. The CamIO app is running on the iPhone 8, held by a gooseneck cell phone holder clamped to the table, with its main camera aimed towards the cell model.

To create a new *hotspot* (location of interest on the object with associated audio label), the user moves the writing stylus tip to the desired location, and dwells there for about a second. CamIO recognizes this dwell gesture and begins recording audio. To stop the recording, the user

moves the writing stylus out of view of the camera, and then the app plays back the audio recording for confirmation. The audio label is triggered whenever the user holds the reading stylus tip within 1 cm of the hotspot. If the user moves the stylus away while the audio label is playing, the announcement is interrupted; this allows the user to halt the announcement before it is complete.

CamIO includes simple audio feedback to provide visibility information to the user: (a) a single-click pulse sounds repeatedly whenever the reading stylus is visible; (b) a double-click pulse sounds repeatedly whenever the writing stylus is visible; (c) a warning sound is issued if the board is not visible to the camera. This feedback helps the user maintain visibility of the styluses and the object at all times. Note that, once the user launches the app, the user need not touch the screen; this is a useful feature since the iPhone is usually rigidly mounted for use with CamIO and in such a configuration it may be difficult for the user to access the touch screen.

Co-Design

We conducted a co-design (Sanders and Stappers) followed by a study with six participants who are blind (4 female, 2 male; age range 26-73, average age 38.5 years). The co-design allowed participants to describe their issues with labeling and create their ideal solution without design fixation of the researchers' solution. In the co-design we asked each participant how they currently use labels (in any form, whether braille, audio or tactile) on objects in their daily life, and whether and how they add their own labels. Next we asked them to imagine an ideal labeling solution, such as what might be available with advanced technology in the future, assuming no limits on what this technology could perform. Finally, we showed them two objects they are likely to use in daily life: a 2D tactile street map and a microwave oven (Fig. 2b,c). For both objects we asked participants how they would label the object today and how they might

label it with a futuristic solution. To analyze participant responses, we used an inductive grounded in vivo coding technique.

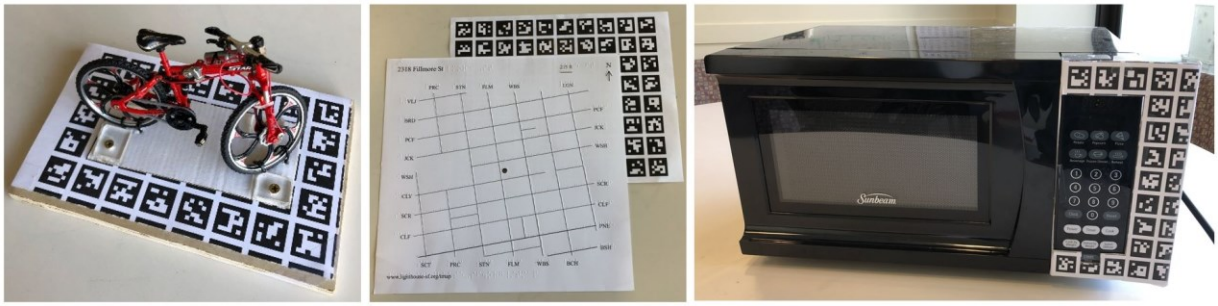

Fig. 2. Remaining objects used in the user study. Left to right: (a) 3D bicycle model rigidly mounted on the board used in the training session. (b) Tactile map (TMAP) of the neighborhood near Smith-Kettlewell in San Francisco, which is attached to the board. (c) Microwave oven, with a board mounted to the side of the touch panel.

The coded responses from the participants during the co-design revealed a number of trends and desired features for CamIO. All participants identified as “lazy labelers,” only labeling what was minimally necessary to use a device with bump dots and minimal braille. The sentiment was if the labels were there already, they would sometimes be useful, but participants did not want to spend the time to label appliances. The ideal labeling solution for all participants included both braille and audio labels. All the participants agreed that existing braille labels were too big, too easily damaged, too expensive, and were too much work for an ideal solution. P5 expressed that “Non-braille labels are only a substitute for braille.” P5 also described a “magic braille” solution that materializes when an object is touched. All participants but P4 wanted something that could accurately recognize objects when held up in front of a camera and apply annotations if created. Particular items that participants wanted labeled were spices, items in the freezer and fridge, cans, and items in their packages. P1, P4 and P6 wanted something that was voice controlled that could identify objects and read labels. Using these designs as a roadmap for

future work, it is important that CamIO can recognize objects in addition to providing the option to display labels in multiple modalities (e.g., audio, visual and refreshable braille display).

User Study

The co-design was followed by the user study, which all six blind participants completed. The study began by introducing the CamIO app running on an iPhone 8 with a brief explanation of how it works, followed by a hands-on training session with a 3D bicycle model (Fig. 2a). In this session, the participants were shown how to trigger pre-defined hotspots and how to define new hotspots, record associated audio labels, and verify the new hotspots. The experimenters offered both verbal guidance and hands-on assistance (for example, moving the stylus to a desired location while the participant held it). Next, each participant was given tasks of progressively increasing difficulty using three objects: a 3D model of a biological plant cell (Fig. 1), a tactile map and a microwave oven. In these tasks, the experimenters positioned the smartphone camera properly and offered verbal assistance to the participants when needed (simulating the kind of help they might receive from a remoted sighted guide service such as Be My Eyes or Aira (Weiss)), but refrained from providing hands-on assistance as was offered in the training demo.

In the biological cell model task, the experimenters verbally guided each participant to touch three cell structures: the nucleolus, mitochondria and chloroplast. The participant was asked to define one hotspot for each structure, then to verify the new hotspots. All six participants successfully recorded the hotspots and verified them, with the following exceptions: P1 was unable to verify one of the three hotspots and gave up (probably because of a slight imprecision in the location of the hotspot); and P3 had to record one hotspot a second time

because of a software bug (which has since been corrected) that intermittently erased hotspots immediately after they were created.

The tactile map used in the next task was an embossed tactile map (TMAP) of the neighborhood (roughly 8 city blocks wide) centered on the Smith-Kettlewell building. The experimenters verbally directed the participant to two locations on the map, the building and a nearby dead end street. They had them manually trace a route from the building to the dead end, add an audio label at each of the four intervening street intersections on this route, and then verify the new hotspots. All six participants successfully recorded the hotspots and verified them, with the following exceptions: one of P1's four hotspots was located in a slightly wrong position, likely because of stylus localization noise; the software bug mentioned above meant that P2 and P3 each had to record one hotspot twice; and stylus localization noise meant that P1 was unable to verify one hotspot and that P5 and P6 each received the wrong announcement at one hotspot.

The final task was to label buttons on a microwave oven panel and press some realistic button sequences. The number buttons 7, 8, 9 and 0 were pre-labeled; each participant was asked to label the Popcorn, Reset and number buttons 1-5. After labeling the buttons the participant was asked to complete three realistic button press sequences by any means they chose: press Popcorn followed by Reset; press 1, 0, 0 followed by Reset; and press 1, 2, 0 followed by Reset. This task was extremely challenging because the microwave panel is a featureless touch screen; however, the microwave makes a beep sound whenever a button is pushed on it, and the experimenter offered the participant the use of bump dots (Borella) to mark salient locations if they wished. Another factor that contributed to the task's difficulty is that the buttons are narrowly spaced on the microwave, which means that stylus tip localization noise made it difficult to reliably define and find the audio labels. Overall, four of the six participants (P1-P4)

were able to create perfect or near-perfect audio labels of the buttons requested; these four were also able to perform the button sequences either perfectly or nearly perfectly, though P3 and P4 used the bump dots instead of CamIO to identify the buttons for the sequences. The other two participants (P5-P6) labeled the requested buttons but these locations were inaccurately placed, either because of user imprecision or stylus tip localization noise, which prevented them from performing the button sequences correctly.

Fig. 3. Study participant using CamIO with Google Cardboard, worn hands-free with a head strap, to access the microwave oven. Participant uses audio feedback from the CamIO app to help aim his head towards the microwave oven panel.

After these tasks, we asked participants P2-P6 to try using (Google Cardboard) with a head strap to hold the smartphone while using CamIO to access the microwave (Fig. 3). The experimenters explained how to use the app feedback to help the user aim the camera properly towards the microwave by moving their head. While P2 was unsuccessful at getting CamIO to work by this method, P3-P6 were all able to read hotspots, and both P3 and P6 were also able to record one or more new hotspots. After using the Google Cardboard, P4 tried holding the

smartphone in one hand while holding the stylus in the other, and was able to read hotspots and create a new one in this way.

We closed the study with a semi-structured interview in which we asked participants to comment on likes and dislikes of CamIO, desired improvements of the system, and on how their conception of an ideal labeling system might change after having experienced CamIO. All participants expressed that they ideally wanted to use their finger rather than a stylus to get a label. All but P4 said one stylus was OK, but did not want two styluses and offered suggestions for removing the second stylus. All participants said they wanted some kind of tactile indication for the location of CamIO labels, whether it be an intrinsic tactile feature or else a braille label or a bump dot on a featureless surface. All participants thought that CamIO would be extremely useful for querying information about an object, such as ingredients, points of interest around the current location, or opening hours. All but P4 said aiming the camera was too difficult and needed to be improved. P1 stated "CamIO with… machine learning [to include object recognition/identification] would be pretty powerful." All participants agreed that CamIO needed work before it could change their ideal labeling solutions, but they could see themselves using it once it was improved.

Conclusions

The new version of CamIO runs as a standalone iPhone app that is controlled by two styluses, which enables exploration of existing hotspots and the accessible creation of new ones. Ongoing work includes periodic user testing and the user interface enhancements to support accessible aiming of the smartphone camera towards the object of interest, culminating in the release of CamIO as a free iOS app on the App Store. Future work will also explore the development of more accurate stylus localization algorithms, the use of a computer vision

algorithm that eliminates the need for the board, the use of entire hotspot regions (such as a lake on a tactile map) in addition to hotspot points, and the use of a single stylus for both reading and writing. Finally, we will explore multiple label display options, including text-to-speech and refreshable braille, and a zoomed-up visual display for low vision users.

Acknowledgments

This work was supported by NIH grant 5R01EY025332 and NIDILRR grant 90RE5024-01-00.

Works Cited

Borella, Megan. "LightHouse Life Hacks: 7 Ways the Bump Dot Can Make Your Life Easier." https://lighthouse-sf.org/2017/07/26/better-living-through-stickers-7-ways-to-use-bump-dots-in-daily-life/ . Accessed 13 Oct, 2019.

Coughlan, James M., and Miele, J. "Evaluating Author and User Experience for an Audio-Haptic System for Annotation of Physical Models." 19th International ACM SIGACCESS Conference on Computers & Accessibility. ACM, 2017.

Fusco, Giovanni and Morash, V.S. "The tactile graphics helper: providing audio clarification for tactile graphics using machine vision." 17th International ACM SIGACCESS Conference on Computers & Accessibility. ACM, 2015.

"Google Cardboard: Experience virtual reality in a simple, fun, and affordable way." https://arvr.google.com/cardboard/ . Accessed 13 Oct, 2019.

Kendrick, Deborah. "PenFriend and Touch Memo: A Comparison of Labeling Tools." Access World Magazine 12.9 (2011).

Mascetti, Sergio, et al. "JustPoint: Identifying Colors with a Natural User Interface." 19th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, 2017.

Miele, J. "Talking Tactile Apps for the Pulse Pen: STEM Binder." 25th Annual International Technology & Persons with Disabilities Conference (CSUN). 2010.

"Oculus Quest: Our first all-in-one gaming headset." Facebook Technologies, LLC. https://www.oculus.com/quest/ . Accessed 13 Oct, 2019.

"OpenCV." https://opencv.org/ . Accessed 13 Oct, 2019.

Sanders, Elizabeth B-N., and Pieter Jan Stappers. “Probes, toolkits and prototypes: three approaches to making in codesigning.” CoDesign 10.1 (2014): 5-14.

Shen, Huiying, et al. “CamIO: a 3D computer vision system enabling audio/haptic interaction with physical objects by blind users.” 15th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, 2013.

Shi, Lei, Yuhang Zhao, and Shiri Azenkot. “Markit and Talkit: a low-barrier toolkit to augment 3D printed models with audio annotations.” 30th Annual ACM Symposium on User Interface Software and Technology. ACM, 2017.

Szeliski, R. **Computer vision: algorithms and applications**. Springer Science & Business Media. 2010.

“Talking Tactile Tablet.” Touch Graphics, Inc. http://touchgraphics.com/portfolio/ttt/ . Accessed 13 Oct, 2019.

Thévin, Lauren, and Anke M. Brock. “Augmented Reality for People with Visual Impairments: Designing and Creating Audio-Tactile Content from Existing Objects.” International Conference on Computers Helping People with Special Needs. Springer, Cham, 2018.

“TMAP: Tactile Maps Automated Production.” LightHouse for the Blind and Visually Impaired. https://lighthouse-sf.org/tmap/ . Accessed 13 Oct, 2019.

Weiss, Martin, et al. “A Survey of Mobile Computing for the Visually Impaired.” arXiv preprint arXiv:1811.10120 (2018).