CamIO in the Browser: A Cross-Platform Audio Label Tool for Tactile Graphics

Dragan Ahmetovic
Dept. of Computer Science
Università degli studi di Milano
Milano, Italy
dragan.ahmetovic@unimi.it

James Coughlan
Smith-Kettlewell Eye Research Institute
San Francisco, California, USA
coughlan@ski.org

Giorgio Dal Santo
Dept. of Computer Science
Università degli Studi di Milano
Milano, Italy
giorgio.dalsanto@studenti.unimi.it

Khadija Ezrouri
Dept. of Computer Science
Università degli Studi di Milano
Milano, Italy
khadija.ezrouri@studenti.unimi.it

Matteo Manzoni
Dept. of Computer Science
Università degli Studi di Milano
Milano, Italy
matteo.manzoni@unimi.it

Sergio Mascetti
Dept. of Computer Science
Università degli Studi di Milano
Milano, Italy
sergio.mascetti@unimi.it

Abstract

For blind or low vision individuals, tactile graphics (TGs) provide essential spatial information but are restricted in the amount of data they can convey – especially semantic information, which is represented using the small amount of braille abbreviations that fit on the TG. Using computer vision, TGs can be enhanced by adding audio labels, in which audio descriptions are triggered when the user touches elements on the TG; these audio descriptions can contain unlimited amounts of semantic information and are accessible even to those who don't read braille. Unfortunately, existing audio label systems are closed systems with severe limitations, some of which are costly and/or tied to specific hardware platforms. We address this problem by creating CamIO-Web, an open-source version of our CamIO (short for "Camera Input-Output") audio label system, which runs in the browser of virtually any computer or mobile device. The system includes facilities that allow users to create their own TGs with any audio labels (audio recordings or Text-to-Speech), and the open-source code base makes it extensible.

CCS Concepts

• **Human-centered computing** → *Accessibility systems and tools*;
• **Applied computing** → *Arts and humanities.*

Keywords

Art accessibility; Audio-tactile interfaces; Visual impairment.

ACM Reference Format:
Dragan Ahmetovic, James Coughlan, Giorgio Dal Santo, Khadija Ezrouri, Matteo Manzoni, and Sergio Mascetti. 2025. CamIO in the Browser: A Cross-Platform Audio Label Tool for Tactile Graphics. In *Adjunct Proceedings of the 27th International Conference on Mobile Human-Computer Interaction (MobileHCI '25 Adjunct), September 22–25, 2025, Sharm El-Sheikh, Egypt*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3737821.3748528

MobileHCI '25 Adjunct, Sharm El-Sheikh, Egypt

ACM ISBN 979-8-4007-1970-7/25/09
https://doi.org/10.1145/3737821.3748528

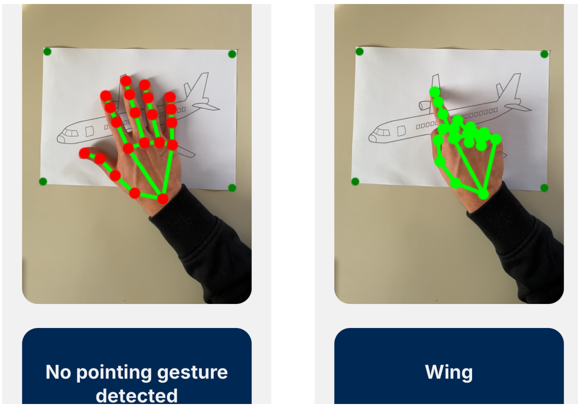
1 Introduction

For individuals who are blind or visually impaired (BVI), tactile graphics (TGs) and 3D models are invaluable resources for accessing information [9]. Materials such as tactile maps, relief maps, and models representing structures like molecules, skeletons, or biological cells can convey spatial and structural information through touch. However, other aspects, including color, visual texture, and printed content, typically remain inaccessible to BVI users. While braille is often employed to label relevant features, it is generally restricted to abbreviated forms due to limited space, requiring separate keys for interpretation. Furthermore, a significant portion of the BVI population does not read braille [7].

To overcome these limitations, existing solutions propose to augment tactile materials with touch based interaction and audio feedback [1]. One such system is the T3 Tactile Tablet[1], which enables users to place a tactile overlay on the tablet's touchscreen. The touchscreen can then detect finger contact with the raised elements of the overlay, providing audio feedback corresponding to the specific tactile feature being explored. This paradigm allows BVI users to access rich semantic content without relying on braille, leveraging a familiar interaction modality that is similar to mainstream touchscreen technologies.

An alternative technical solution to achieve similar functions is to use computer vision to transform virtually any TG or 3D model into an interactive interface, without being limited to thin, two-dimensional touchscreen overlays. Prior research in this area has explored a range of solutions, including depth-sensing cameras [4, 6], conventional webcams [5], and smartphone-based applications [2, 3, 10]. Some approaches require the user to interact with tactile materials through a dedicated pointing device [2] or to make their fingertip more visible with a visual marker [5]. More recently, advances in hand tracking technologies have enabled the development of solutions based on hand pointing and gestures without the need for pointing devices[2,3] [6, 10]. These interfaces are particularly well-suited to the exploration of tactile materials, as they accommodate the natural use of multiple fingers and both hands.

[1] https://www.touchgraphics.com/store/t3-t3-books
[2] https://tactileimages.org/en/reader-app/
[3] https://www.tactonom.com/en/tactonome-reader/



(a) Open hand. (b) Pointing gesture.

Figure 1: CamIO-Explorer interface on a mobile device

Following this line of research and development, we demonstrate CamIO-Web, a web-based, cross-platform, open-source[4] application that supports the labeling and afterwards the touch exploration of TGs providing the inserted labels as audio feedback. Specifically, CamIO-Web consists of two main components: CamIO-Explorer and CamIO-Creator. A user can access CamIO-Creator to create audio labels for a given tactile model using an intuitive web interface. Then CamIO-Explorer can be used, on mobile devices or a computer, to support the exploration of the tactile model by reading aloud the labels defined with CamIO-Creator. While our focus is on BVI users, we note that CamIO-Web may also be useful for sighted people who benefit from multi-modal interactivity, which combines the visual, audio, and tactile senses.

2 The CamIO-Web system

The CamIO-Web system offers the convenience of an audio label system that runs entirely in the browser, which has the advantage that it runs on virtually any computer or mobile device. Moreover, the system runs without requiring the user to install any software – the user only needs to click on the CamIO-Web link to run it. Finally, since the software runs entirely on the client, it is highly responsive, eliminating the latency that would be incurred if camera data had to be transmitted to a remote server to be processed and the results sent back to the client.

[4]Code is available at https://gitlab.di.unimi.it/ewlab/public/camio/camio-web

2.1 Architecture and data structures

CamIO-Web is a public website[5], composed of two main web apps: CamIO-Explorer and CamIO-Creator. Currently, the server has the only purpose of hosting the website and it does not store user-generated content. To allow the user to store the models descriptions communication between the two apps, we defined the `.camio` file format that contains the following information:

- Template, an image of the TG.
- List of hotspot areas, each representing a specific region of the TG and associated with a title, a color, and, optionally, a textual description and an audio file.
- Color map: an image file used to encode the position of the hotspot areas used to represent the position of each hotspot area through its associated color.

2.2 CamIO-Explorer

CamIO-Explorer provides audio feedback while the user explores the TG. The user can freely explore the TG with an open hand, without triggering feedback from CamIO-Explorer (Figure 1a). Instead, CamIO-Explorer describes the touched hotspot area when it detects a pointing gesture (Fig. 1b), which is defined as a gesture in which the index finger of one hand (left or right) is pointed straight while the other fingers are bent or closed. The description consists of an optional sound and a text string, which can be either shown on the web interface or read aloud with the built-in speech synthesizer. Note that CamIO-Explorer is designed to be compatible with the system screen reader, thus, upon updating the text label, the screen reader reads the corresponding text.

Figure 2 shows the computational pipeline adopted by CamIO-Explorer. CamIO-Explorer acquires the image from the camera feed and runs a template matching algorithm to map the model template to where it appears in the image. Since the operation is computationally intensive, it is not run for every frame, but every 2 seconds. For every frame, instead, CamIO-Explorer uses the MediaPipe library[6] to extract the hand skeleton model, which is then processed with a rule-based procedure to determine whether the hand is making the pointing gesture (a "number one" gesture with the hand making a fist and the index finger pointing straight). If so, CamIO-Explorer checks if the pointing position is within a hotspot area. If so, the feedback associated to the hotspot is provided to the user.

[5]https://develop.ewlab.di.unimi.it/camio
[6]https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

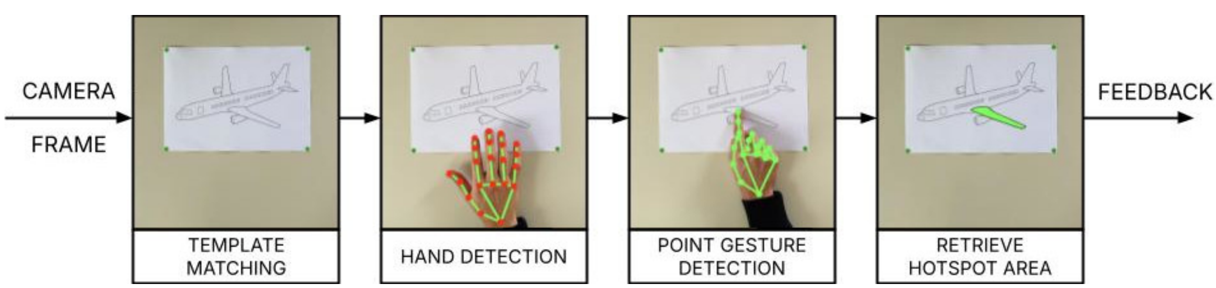


Figure 2: CamIO-Explorer pipeline

2.3 CamIO-Creator

Figure 3 shows the main interface of CamIO-Creator. There are three main graphical elements. On the left, the application presents the list of hotspots. The user can create a new one or edit an existing one (see Figure 4a). Note that the the specific color chosen is irrelevant: its only purpose is to clearly distinguish one hotspot from the others. In the top part of the interface, CamIO-Creator shows the toolbar, reminiscent of basic image editors. From this bar, the user can select how to draw on the image. The center of the screen shows the template image where the user can draw to create the color map. Upon saving, the user can specify the model's name and two descriptions (Figure 4b).

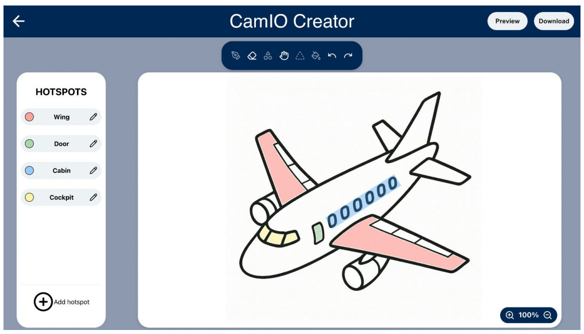
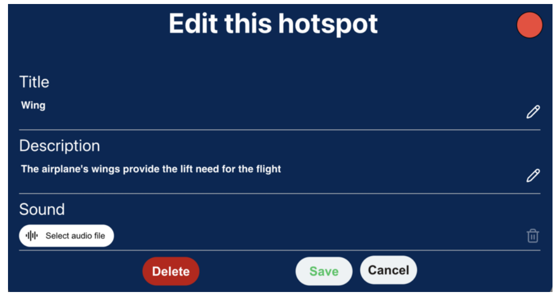
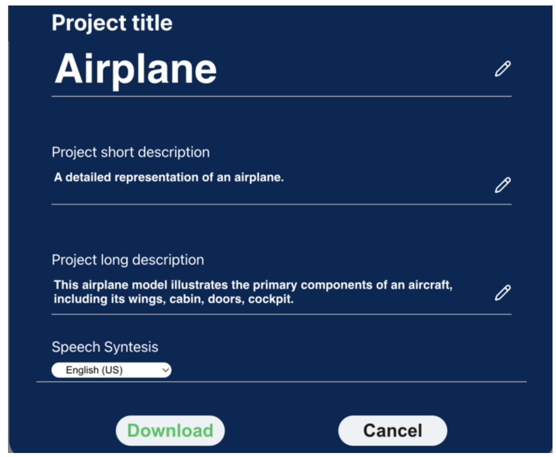


Figure 3: CamIO-Creator main interface.



(a) Hotspot: create or edit an hotspot



(b) Saving a model with name and descriptions.

Figure 4: CamIO-Creator interfaces

3 Conclusions and limitations

The demo will showcase CamIO-Web and highlight its value for end-users and for researchers. For end-users, CamIO-Web is a public application already available to support the exploration of TGs by BVI people, with a number of practical advantages: it is free, does not require installation of software, is compatible with most system screen readers, and can be used on traditional and mobile devices. The fact that CamIO-Explorer runs on mobile devices is particularly convenient for students, who do not need to transport bulky equipment. Second, it is a technological platform that can be easily extended to experiment with new forms of interaction. In particular, we are currently investigating the use of a larger set of hand gestures, the integration of CamIO-Explorer with Large Language Models, and the user of wearable devices, like Ray-Ban Meta[7] [8]. While CamIO-Web currently supports models that are flat (2D) or nearly so, in the future we will extend the system to work with 3D models.

Acknowledgments

James M. Coughlan and Sergio Mascetti were supported by the National Eye Institute/National Institutes of Health under Grant 2R01EY025332; James M. Coughlan was also supported by the National Institute on Disability, Independent Living, and Rehabilitation Research under Grant 90REGE0018 and The Smith-Kettlewell Eye Research Institute.

[7] https://www.meta.com

References

[1] Anke M Brock, Philippe Truillet, Bernard Oriola, Delphine Picard, and Christophe Jouffrais. 2015. Interactivity improves usability of geographic maps for visually impaired people. *Human–Computer Interaction* (2015).

[2] James M Coughlan, Huiying Shen, and Brandon Biggs. 2020. Towards accessible audio labeling of 3D objects. In *Journal on technology and persons with disabilities: Annual International Technology and Persons with Disabilities Conference*. CSUN.

[3] Giovanni Fusco and Valerie S Morash. 2015. The tactile graphics helper: providing audio clarification for tactile graphics using machine vision. In *International ACM SIGACCESS Conference on Computers & Accessibility*. ACM.

[4] Huiying Shen, Owen Edwards, Joshua Miele, and James M Coughlan. 2013. CamIO: a 3D computer vision system enabling audio/haptic interaction with physical objects by blind users. In *International ACM SIGACCESS Conference on Computers & Accessibility*. ACM.

[5] Lei Shi, Yuhang Zhao, and Shiri Azenkot. 2017. Markit and Talkit: a low-barrier toolkit to augment 3D printed models with audio annotations. In *Annual ACM symposium on User Interface Software and Technology*. ACM.

[6] Xiyue Wang, Seita Kayukawa, Hironobu Takagi, and Chieko Asakawa. 2023. TouchPilot: Designing a Guidance System that Assists Blind People in Learning Complex 3D Structures. In *International ACM SIGACCESS Conference on Computers and Accessibility*. ACM.

[7] Jaroslaw Wiazowski. 2014. Can Braille be revived? A possible impact of high-end Braille and mainstream technology on the revival of tactile literacy medium. *Assistive Technology* (2014).

[8] Nasif Zaman, Venkatesh Potluri, Brandon Biggs, and James M Coughlan. 2025. WhatsAI: Transforming Meta Ray-Bans into an Extensible Generative AI Platform for Accessibility. *arXiv preprint arXiv:2505.09823* (2025).

[9] Kim T Zebehazy and Adam P Wilton. 2014. Straight from the source: Perceptions of students with visual impairments about graphic use. *Journal of Visual Impairment & Blindness* (2014).

[10] Maralbek Zeinullin and Marion Hersh. 2022. Tactile audio responsive intelligent system. *IEEE Access* (2022).