

The Haptic Video Player: Using Mobile Robots to Create Tangible Video Annotations

Darren Guinness, Annika Muehlbradt, Daniel Szafir, Shaun K. Kane

University of Colorado Boulder

Boulder CO, USA

{darren.guinness, annika.muehlbradt, daniel.szafir, shaun.kane}@colorado.edu

ABSTRACT

Video and animation are common ways of delivering concepts that cannot be easily communicated through text. This visual information is often inaccessible to blind and visually impaired persons, and alternative representations such as Braille and audio may leave out important details. Audio-haptic displays allow for the presentation of complex spatial information, along with accompanying description. We introduce the Haptic Video Player, a system for authoring and presenting audio-haptic content from videos. The Haptic Video Player presents video using mobile robots that can be touched as they move over a touch screen. We describe the design of the Haptic Video Player system, and present user studies with educators and blind individuals that demonstrate the ability of this system to render dynamic visual content non-visually.

Author Keywords

Accessibility; Tangible User Interfaces; Education; Video; Robots; Blindness.

CCS Concepts

Human-centered computing → Accessibility systems and tools.

INTRODUCTION

The ubiquity of camera-enabled devices, video editing software, and free online hosting has resulted in video becoming a primary method for sharing information online. Video can be especially useful in communicating spatial information, such as demonstrating the movements of objects in the solar system or sharing an exciting play in a sports game. Increasingly, video content is used in educational contexts, and students around the world can learn independently from sites such as Khan Academy and Coursera. Representing content as video or animation can result in improved learning outcomes for some topics [17].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ISS '18, November 25–28, 2018, Tokyo, Japan

© 2018 Copyright is held by the owner/author(s).

Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5694-7/18/11 \$15.00

<https://doi.org/10.1145/3279778.3279805>

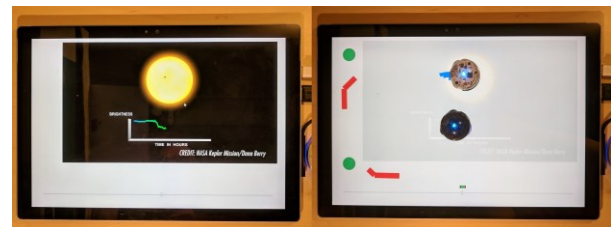


Figure 1. Side-by-side view of the original video of a small satellite crossing in front of a star and a graph of a temporarily decreasing brightness level (left), and a haptic representation of the same frame with robots representing the satellite (top robot), and the relative brightness level on the graph (bottom robot).

Most video content is inaccessible to visually impaired persons (VIPs). While videos can be made more accessible by providing a descriptive audio track, a verbal description may not fully convey the content of a video. For example, consider an educational video that demonstrates gravity and acceleration by showing a cannonball as it is fired. An audio description of the video may describe the overall video, but may not be effective at conveying the shape of the cannonball's curve, or the relationship between its horizontal speed and vertical speed. Conversely, a tactile graphic might convey the arc of the cannonball, but cannot capture subtle changes in speed over the course of the trajectory. For representing dynamic visual content such as video and animation, the medium itself should convey motion.

Recent work in dynamic shape displays raises the possibility of animated tactile representations of visual information [7,10,15,20,24]. Shape displays offer the potential to convey spatial relationships and movement to VIPs through tactile sensations [24]. However, these displays often present usability trade-offs, requiring significant instrumentation of the environment [7,8,15,24]. Furthermore, many of the publicly known shape displays exist only as research prototypes and are unavailable to consumers. There exists an opportunity to use off-the-shelf technologies to create shape displays that can be accessed by anyone.

To address these issues, we introduce the Haptic Video Player, a system that uses off-the-shelf technology to annotate existing video content with tangible data, and plays this video back using a mobile robot-based shape display. We provide insights from interviews with members of the

community about accessibility barriers in online video. We also demonstrate the platform’s ability to convey shapes and spatial relationships in an evaluation with 7 VIPs.

While providing alternate forms of video is essential, a viable solution must also provide a streamlined way of authoring accessible video content. We support authoring of accessible video annotations through an authoring tool that allows a content author to use on-screen gestures to direct the movement of the robots, which is then synced to the video.

The contributions of this paper are:

1. The Haptic Video Player, a software system which enables direct tangible reflection of spatial events in videos for individuals with vision impairments;
2. The Haptic Video Authoring Layer, a software package which allows a user to create haptic videos overtop of existing video content;
3. Insights from interviews with members of the blind and visually impaired community about inaccessible video content;
4. Insights from a user study in which 7 VIPs used the display to infer spatial relationships in online video content.

RELATED WORK

Making Video Content Accessible

Currently, most video is made accessible by supplementing it with audio description. Inaccessible video can also be represented as a tactile graphic or through sonification.

Audio description: A sighted author can provide narration that describes what is happening in a video. This audio can be played along with the video and its original audio track.

While researchers have often explored ways to effectively create descriptions of static images (e.g., [9,16,18,22,30]), relatively little work has explored how to create or render audio descriptions for video. YouDescribe [31] is a web application that simplifies the process of recording, finding, and playing back audio descriptions. A volunteer can load any video hosted on YouTube and can record a description for that video. A blind web user who is browsing YouTube can then browse available descriptions for a video. The YouDescribe player automatically integrates the description into video playback by pausing the video when the description is being read. Our work adopts a similar approach in enabling accessible information to be added to a video, but supports both audio and tactile annotations.

Tactile graphics: Tactile graphics can be useful in presenting spatial information in an accessible format [6]. Tactile graphics can be used to represent maps [2,25], data graphics [12], picture books [23], and other images. Tactile graphics can also be annotated with audio descriptions [4]. Tactile graphics typically focus on representing a static image, and cannot easily represent moving images.

Sonification: Moving images can be made accessible by developing a scheme for sonifying that video. For example, a video sonification might track objects in a video and represent their visual characteristics, such as color and movement speed, through sound parameters, such as pitch, volume, or stereo position. Sonification has been used to create accessible representations of aquaria [11,28] and simulations of our solar system [26]. Sonification can provide additional access to moving images, but is limited in the amount of information that can be conveyed at once.

Tactile and Audio-Tactile Displays

Representing information in a tactile format requires developing or using a tactile display, as most mainstream computing devices do not provide sufficient tactile resolution to convey images. These techniques use a variety of techniques to represent tangible information, including vibration [14], ultrasonic haptics [29], electrostatic vibration [1], pin arrays [7, 27], haptic styli [21], and electromagnetic markers [24]. These systems tend to use specialized hardware, which may be expensive or may be available as a research prototype only.

More recently, researchers have used mobile robot swarms to create tactile displays. These systems have the benefit of being modular so that robots can be added or removed. Examples of this approach include Zooids [15], BitDrones [8], Touchbugs [19], Cellulo [20], and Tangible Reels [5]. Most of these projects either require extensive augmentation of a room with sensors [8, 15] or require custom hardware which can be difficult to obtain or costly [5, 20, 24] limiting their adoption and potential applications. Other projects, such as GUI Robots [10], use off-the-shelf robots, enabling the creation of tactile displays without the need for any custom hardware. These robot-based displays can present tactile information by moving the robots over a work surface to draw shapes or indicate direction. The Haptic Video Player uses off-the-shelf robots to create a tactile display, and provides tools for authoring and viewing audio-tactile video content.

THE HAPTIC VIDEO PLAYER

We present the Haptic Video Player, a software-hardware system that allows users to author and consume *haptic annotations* to existing videos. In a haptic annotation, one or more small wireless robots reflect the motion of important spatial events or relationships which can be understood by individuals with vision impairments through touch.

Development of the system has been guided by the following design goals: (1) **Ease of authoring:** Non-expert users must be able to create accessible video content from existing videos; (2) **Versatility:** The system should be able to support a variety of existing videos and animations; (3) **Fidelity:** The accessible videos should accurately represent the source material, correctly rendering the location and motion of on-screen objects. (4) **Understandability:** The system should enable a VIP to explore and understand the source content.

System Design

The Haptic Video Player combines off-the-shelf hardware (a touch screen tablet and a set of mobile robots) with custom software for creating and rendering accessible video content.

Hardware: The current prototype uses a Microsoft Surface Pro 4 tablet (12.3-inch, 3:4 ratio touch screen), but should work on any Windows 10 tablet. Tangible feedback is provided by a set of Ozobot Bit robots (Fig. 2). These robots are approximately one cubic inch in volume, and feature two wheels and an optical sensor for guidance. We chose these robots because of their versatility, ergonomic size, and low cost (about \$50 USD).

Users are provided with several methods for exploring videos. The user may explore the video by touching the robots or screen, and may navigate through videos using an attached QWERTY keyboard or a Griffin PowerMate knob.

Software: We developed two applications to enable the creation and consumption of accessible videos. First, we developed an authoring tool, *HVAuthor*, which enables a sighted author to annotate an existing video with audio and tactile information. The corresponding player, *HVPlayer*, allows a blind person to explore an annotated video. *HVPlayer* uses a framework that we have developed internally, *RoboKit*, to control the robot devices. These components are described in further detail below.

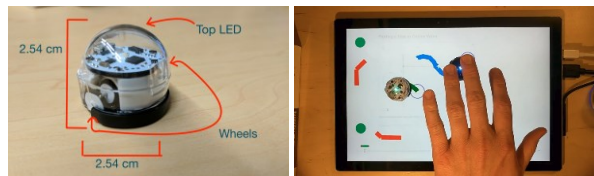


Figure 2. (left) The Ozobot robot used in the studies. (right) User tracking a cosine wave drawn during a video.

Representing Motion via Robots

The Haptic Video Player presents tangible annotations within a video by controlling the movements of small robots over the screen (Fig. 2). This approach allows us to create a tangible display using off-the-shelf hardware, and allows for displaying multiple objects that move simultaneously.

After exploring several robot platforms, we chose the Ozobot Bit robot for our prototype. This robot is controlled via an optical sensor on the bottom of the robot. By default, the robot follows lines beneath it; lines of different colors affect the speed. The robot also responds to a set of predefined optical codes that show a sequence of flashing colors. These codes can instruct the robot to stop, change speed, or turn around. One advantage of this control scheme is that the robot can be controlled without the use of an external tracking camera, which could be occluded by users' hands.

To adapt these robots for use in our system, we developed a framework, *RoboKit*, that dynamically draws lines and color codes on the screen to control the robots. This framework

allows the system to direct a robot to a specified point, or to follow a specified path.

Although the robots can switch between several preset speeds (25, 45, 65, 85 mm/sec), matching the movements of a video (e.g. object A is moving at 60% of object B) may require fine-grained control of speed. To manage the robots' speed, *RoboKit* draws lines in a flashing pattern to interpolate between the preset speeds when necessary.

Haptic Video Annotations and Lifecycle

While our approach uses robots to represent movements on screen, this approach can break down in some situations, such as when the camera changes in a video and objects jump from one location to another. Thus, rather than having tactile representations of an entire video, we allow authors to annotate multiple haptic segments within a single video. For example, a video recording of a lecture on kinematics might mostly consist of a talking head, interspersed with animated demonstrations. In this case, we may want to provide tangible feedback during the demonstration, but not for the talking-head segments.

When the user plays back a video, the video is presented as normal. When playback reaches an annotation, the video pauses as the robots are set up. The system plays a sound to indicate that the annotation is beginning and then reads an annotator-created scene description as the robots move into place. This description helps the user understand the context of the original scene and includes details such as the camera's position, the objects in the scene, and an overview of the events and activity to be reflected in the annotation. This process is shown in Figure 3.

After the annotation is loaded, each robot proceeds to reflect or mirror a virtual object's motion on top of the display. The user can track the motion of the robot with their hands to learn the position of the corresponding on-screen object. The user has the option to pause, rewind, or restart the haptic video during playback.

After the haptic annotation is finished, the annotation mode ends. The system announces that the annotation has ended, and the robots move back to their idle position at the corner of the screen.

Authoring Audio-Tactile Video Content

Our system enables non-expert users to add accessibility annotations to existing video content. While these annotations could be added by the original video creator, this method of annotation means that accessible videos may be produced by volunteers, teachers, or others, even if they do not have the ability to edit the video source.

To achieve this goal, we developed the Haptic Video Author, or *HVAuthor* (Fig. 4). *HVAuthor* is used to add haptic annotation segments to existing videos. Each segment consists of a scene description and a series of robot paths.

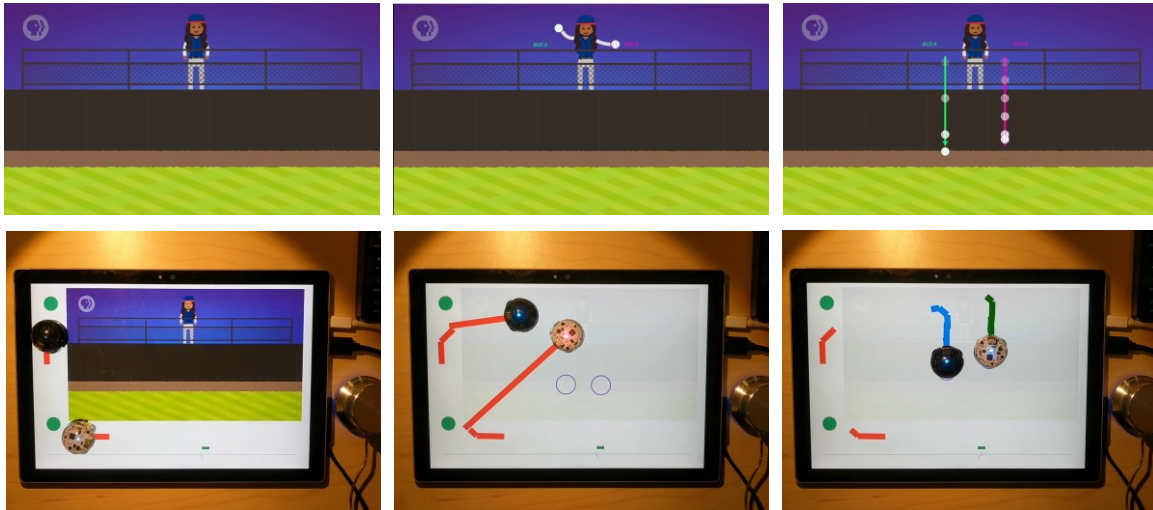


Figure 3. Haptic video playback. The original video is shown above, while the haptic video representation is shown below. These snapshots represent various points in lifecycle of a haptic video annotation: (1) two robots sit idle waiting during unannotated portions of the video; (2) an annotation begins, the robots move to their initial positions, and the system provides a spoken description of the scene; (3) the robots move downward, reflecting the motion of the on-screen falling objects.

Creating an annotation segment involves the following steps:

1. The author loads a video into HVAuthor. Currently, the author can choose a local video file or can enter a URL for a video on YouTube;
2. The author scrubs through the video and marks which segments will have accessible annotations;
3. For each segment, the author can add metadata describing that segment, including a description of what the segment is showing and a list of the objects that will be rendered as tangible objects. This metadata can be added as text or as a voice recording;
4. For each segment, the author adds robots to the video by touching the screen and dragging their finger along the robot's intended path. The user can scrub through the video for precise authoring or play the video in real-time as the author draws a path.

Currently, HVAuthor is optimized for adding a single robot at a time. If the video involves multiple robots, as in Figure 3, the author can one robot, rewinds the video, and add the second robot, or add multiple robot paths through multitouch. The user can also preview the haptic animation at any time; they may correct a mistake by rewinding the video and redrawing the robot path.

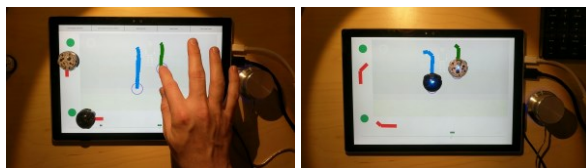


Figure 4. The Haptic Video Author. (left) An author annotates a video by drawing robot paths as the video plays. (right) During playback, the robots move along the specified paths to represent on-screen objects.

Note that it is possible to author robot paths in HVAuthor that could not be played back using the current prototype system. For example, an author might draw a path that moves faster than the robot's maximum speed or draw two robot paths such that the robots would crash into each other. HVAuthor will detect these problems and notify the author although future versions of the software could address this problem in other ways, such as by slowing the speed of the video playback so that the robots could keep up, or automatically rerouting the robots to avoid collisions.

Currently, HVAuthor requires a sighted person to adapt the on-screen content of a video into a haptic annotation. We discuss opportunities to adapt this process so that it can be completed by a blind person in the future work section.

Interacting with Audio-Tactile Video Content

The HVPlayer application (Fig. 3) is used to play back annotated videos. When a user loads an annotated video file, HVPlayer begins playing the source video. The user can watch the video at its normal speed or can use the keyboard's arrow keys to skip between annotated video segments.

When the player reaches the first annotated video segment, it announces the name and description of that segment. If an inactive robot is on the screen, that robot will move to the starting position for the segment. If no robots are available, the system will prompt the user to add a robot; the user places a robot on the screen, next to a tactile marker on the left edge of the screen, and the robot moves to its starting position. As each robot is added, RoboKit draws a path from the starting. Once all of the robots are placed, the user presses a key to start the segment. During the annotated segment, any additional audio description provided by the author is read aloud the robots move along their specified paths.

During each segment, the user can control the video playback using the arrow keys, or by rotating the hardware knob control. During playback, touching anywhere on the screen will read out the name of the closest robot. In addition, the user can double tap the screen to hear the names and approximate locations of all on-screen robots using conventions for accessible talking touch screens [13].

DESIGN INTERVIEWS AND USER TESTING

The Haptic Video Player has been developed through a process of iterative design and evaluation with end user groups, including formative interviews with blind technology experts, and user testing of system prototypes.

Expert Interviews on Accessible Video Content

To identify key content and features for our prototype system, we conducted interviews with three accessibility experts about their current practices surrounding inaccessible video content. All three of our informants were blind screen reader users and worked in professional positions related to accessible technology. All three informants were male. At the start of the interview, we explained that our goal was to develop technology to improve the accessibility of video content. We did not demonstrate prototypes in these sessions, but instead asked informants to share their experiences as both consumers and producers of accessible content.

Encounters with Inaccessible Video

We asked informants about their experiences with inaccessible video content. Informants often consumed online video, including educational and instructional videos, and entertainment such as videos and movies.

The informants mentioned that they sometimes asked a sighted person for help with a video, and tended to ask friends or coworkers, echoing prior research [1]. However, informants tended to avoid asking for help unless it was necessary, as they did not want to be a burden on others [3].

Our informants developed strategies to understand what was going on in a video, even if no description was provided. One informant mentioned that he practiced identifying sound cues in videos to more easily understand what was going on without help.

When they did ask for help, our informants tended to ask people with prior experience describing video to a blind person, so that they would know what to focus on and what to ignore in their description:

I have a friend who describes things to me. We have a built-up vocabulary. I need salient details. We don't bother with verbs. I need concise in real-time. It's hard to do and only very few people can narrate a movie for me ... while they are watching it.

This same informant mentioned that knowing the correct items to describe often required knowing the context and purpose of the video content.

Requirements for Accessible Video Technology

We asked our informants about the kinds of video content that they would be interested in accessing and about how

they would prefer to access that information. Informants pointed out that the vast majority of online video content is not accessible to individuals with vision impairments. This lack of accessibility is most important in contexts related to education, and on large video platforms such as YouTube.

When asked how they would like to interact with accessible video, informants stated that they preferred both tactile and audio representations, depending on the context. Informants noted that tactile landmarks and shapes could be especially useful when exploring videos that involve motion or other spatial information.

When asked about their desires for new technology to support access to video content, informants described the importance of providing additional information without clashing with existing audio descriptions and noted that tactile information is rarely provided along with video. Informants also noted the importance of providing access to existing video content on large platforms such as YouTube. Our Haptic Video Player addresses both comments by providing a way to add additional description and tactile information to existing online videos.

Evaluating the Haptic Video Display

Our goal in developing the Haptic Video Player was to create a system that would provide access to rich information about the activities taking place in a video, including the location and motion characteristics of moving objects. To evaluate the effectiveness of our system, we conducted an evaluation study in which seven blind participants (separate from our interview informants) viewed various types of content using our prototype and described what they perceived.

Participants

We recruited 7 participants (4 male, 3 female, age 22–56) via university mailing lists and via local blindness organizations. Participants' level of vision varied: P1 described his vision as "just above legally blind"; P4, P6, and P7 reported some light perception; P2, P3, and P5 reported that they had no vision. Participants' occupations included university student, teacher, artist, author, and accessibility specialist.

Procedure

Each study session took approximately 45 minutes. After completing the consent process, the researcher introduced the participant to the tablet, robots, and hardware controls. Participants were then shown how to control video playback and how to locate objects on the touch screen.

To evaluate our prototype's ability to convey static and dynamic visual information, we showed participants a series of five demonstration videos, each demonstrating a specific visual feature that could be rendered by the Haptic Video Player and asked the participant to describe what was happening in the video. The study tasks were presented in a fixed order as they increased in complexity videos built upon concepts from earlier tasks. Demo videos included both audio description and haptic annotations. The description of each video and associated task is detailed below:

P	Task 1	Task 2	Task 3	Task 4	Task 5
1	Correct	<i>Reported oval</i>	Correct	<i>Incorrect direction</i>	Correct
2	<i>Incorrect location</i>	Correct	Correct	Correct	Correct
3	Correct	Correct	Correct	Correct	Correct
4	Correct	<i>Reported oval</i>	Correct	Correct	Correct
5	Correct	Correct	Correct	Correct	Correct
6	Correct	<i>Reported oval</i>	<i>Bot names swapped</i>	<i>Incorrect direction</i>	Correct
7	Correct	Correct	Correct	Correct	<i>Bot names swapped</i>

Table 1. Participant responses to each of the five study tasks, coded by the research team. *Italic text indicates differences between the rendered content and participants' descriptions of that content.*

Task 1 - Locate the robot. The robot drove from the top left screen corner to the bottom right. Participants named the quadrant the robot had moved to (i.e., bottom right).

Task 2 - Describe the shape of a robot's path. The robot drove in a circular motion on the screen. Participants described the shape made by the robot's path (i.e., a circle).

Task 3 - Locate two robots. Two robots were placed on the screen. One robot drove from top left to bottom right, the other drove from bottom right to top left. Participants described the final location of both robots.

Task 4 - Identify the direction of a robot's movement. A single robot drove in a diagonal line from top left to bottom right. Participants described the direction of the robot's movement.

Task 5 - Compare movement speed between robots. Two robots were placed on the screen. Each robot moved back and forth across the screen in a horizontal pattern; one moved at 10mm/sec, and the other moved at 30mm/sec. Participants noted which robot was faster.

For each task, participants indicated their answer verbally. Following these tasks, participants used the Haptic Video Player to view a video, created by the researchers, that showed the moon revolving around the Earth. Participants were given time to freely explore the video and to ask questions about it. Following this demonstration, participants provided feedback about the haptic video prototype and their experience in the study.

STUDY FINDINGS

Video Description Accuracy

We measured whether participants were able to correctly describe the five videos presented in the main study task. Prior to analyzing this data, the research team agreed upon a rubric for evaluating each description. For example, for the first video that showed a circle, the answer would be considered correct if the participant described it as a circle; and would be marked incorrect if the participant described another shape. The first author analyzed the participants' responses, which are summarized in Table 1.

In general, participants were able to correctly describe the content of the demonstration videos. Of the 35 video descriptions, 28 were correct based on our rubric, and 7 responses indicated some misunderstanding of the video content. These misunderstandings were minor: three participants described the circle shape as an oval (which was technically true) or spiral; two participants correctly described the location of two robots, but reversed the names of the robots; and three participants incorrectly described the direction or position of a robot. Although the demonstration videos were simple, found that novice users could usually identify locations, shapes, and movement speeds on the display, even when multiple robots were moving simultaneously.

Manual Interactions with Robots

We examined video recordings of the study to identify how participants touched and interact with the robots; we did not specify how to hold or touch the robots. However, participants seemed to settle on various strategies for interacting with the robots (Fig. 5), including hovering one's hand loosely over the robot, tracing behind the robot with a finger, and gripping the top of the robot.

We noted that participants seemed to have favored strategies for interacting with the robot: three used hovering most often, two used tracing, one used gripping, and the last switched between these approaches. In some cases, the participants' hand blocked the robot from moving or knocked it off track. During the study, the experimenter simply replaced the robot, although future prototypes could enable the user to "recalibrate" the system by removing the robots and placing them back in the starting area.

Although we designed this prototype to provide haptic feedback, the presence of the robots sometimes served as multimodal information: one participant mentioned using his remaining vision to follow the robots' movement, and another mentioned that he tracked the robots' location by listening to the sound of their motors.

Subjective Responses and Suggested Applications

We asked participants to rate how *enjoyable*, *easy to use*, and *informative* they found the prototype to be, using a 5-point Likert scale. Participants generally thought that the prototype

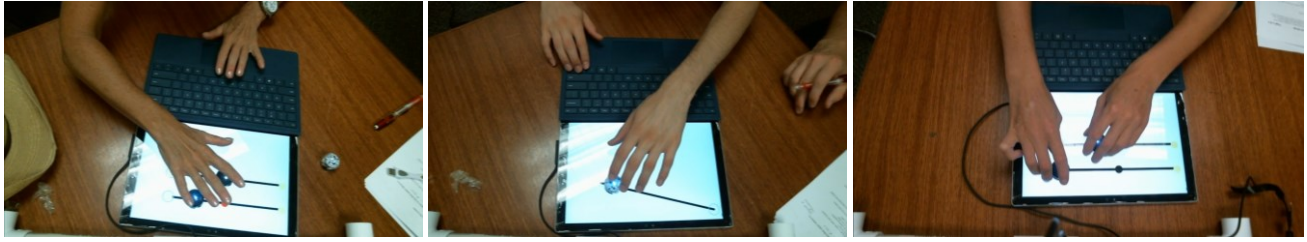


Figure 5. Participants used different poses to touch and follow the robots. From left to right: hover, trace, and grip.

was enjoyable ($M=4.43$, $SD=0.53$) and easy to use ($M=4.57$, $SD=0.54$), but were more reserved in rating the prototype as informative ($M=3.86$, $SD=0.90$). It is possible that the low rating for informativeness is due to the simplicity of the prototype or of the content videos; in either case, these issues should be addressed in future versions of the system.

We asked participants about how they would improve the prototype. Three participants requested a larger display area, and one participant requested that the robots be made smaller. Two participants requested more detailed audio output, such as describing the location of items on screen so that they would be easier to find by touch. Overall, participants expressed interest in the prototype, and most mentioned that they would like to use it in the future.

When asked how they might wish to use this system in the future, participants suggested a variety of uses, including exploring maps and turn-by-turn directions, experiencing sculpture and visual art, representing information on a blackboard or whiteboard, and representing educational books or lessons. Participants mentioned a variety of educational topics that the system might be useful for, such as learning about solar eclipses, knot-tying, and handwriting.

DISCUSSION

Our prototype demonstrates one way in which off-the-shelf technology can be repurposed to serve as a low-cost haptic display. Our user study showed that this type of display can convey spatial information including location, motion, speed, and shape.

Notably, our system is software-only in that all hardware components can be easily purchased and can be used without hardware modification; we thus offer a new opportunity for getting accessible haptic displays into the hands of users.

However, using off-the-shelf robots offers a different set of trade-offs. In particular, while the Ozobot robots are small and inexpensive, their low power motors limit their ability to push or move the user's hand (or other objects on the screen). These robots also have a non-holonomic control system, meaning that they cannot freely move in any direction, and sometimes require space to turn and change direction. Because the robots must steer and make turns to change direction, they cannot always exactly follow the movements of on-screen objects. We expect that future consumer robots will only increase in movement capability, and that we will

be able to include these more advanced robots in future versions of our system software.

One theme that reappeared throughout our user testing was that participants had a variety of preferences for the system hardware: some participants requested a larger display, while others noted that they appreciated the system's portability. As the Haptic Video Player combines multiple pieces of off-the-shelf hardware with custom software, it may be possible to create a version of the Haptic Video Player that supports a variety of sizes, uses, and number of robots, or even a version that enables users to incorporate their own hardware devices.

FUTURE WORK

To date, our evaluation of the Haptic Video Player has focused on verifying the feasibility of our approach and the ability of the system to convey spatial information. In the future, we intend to conduct evaluations in the field in which teachers and students can work together over an extended period of time to co-design accessible video content. We expect that the approach shown here could also be used to provide access to other visual media, such as still images or 3D scenes.

While we have introduced a method that enables end users to add accessible annotations to video, this process currently requires assistance from a sighted person. This barrier should be overcome as soon as possible, and we believe that we will be able to create a fully independent version of the Haptic Video Player in the future, perhaps by using a combination of computer vision and crowdsourcing to enable blind users to independently generate accessible videos.

CONCLUSION

In this paper, we introduce the Haptic Video Player. Our system provides a software framework for adding audio and tactile annotations to existing videos, and allows a blind or visually impaired user to explore this content via a novel audio-tactile display device. Our work to date shows that many online videos can be improved by adding audio and tactile annotations, and that off-the-shelf hardware can be repurposed to enable multimodal, accessible video content.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under grants IIS-1619384 and IIS-1652907. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES

1. Bau, O., Poupyrev, I., Israr, A., & Harrison, C. (2010, October). TeslaTouch: electrovibration for touch surfaces. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology* (pp. 283-292). ACM.
2. Branham, S. M., & Kane, S. K. (2015, April). Collaborative accessibility: How blind and sighted companions co-create accessible home spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2373-2382). ACM.
3. Branham, S. M., & Kane, S. K. (2015, October). The invisible work of accessibility: how blind employees manage accessibility in mixed-ability workplaces. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (pp. 163-171). ACM.
4. Brulé, E., Bailly, G., Brock, A., Valentin, F., Denis, G., & Jouffrais, C. (2016, May). MapSense: multi-sensory interactive maps for children living with visual impairments. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 445-457). ACM.
5. Ducasse, J., Macé, M. J., Serrano, M., & Jouffrais, C. (2016, May). Tangible Reels: construction and exploration of tangible maps by visually impaired users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2186-2197). ACM.
6. Edman, P. (1992). *Tactile Graphics*. American Foundation for the Blind.
7. Follmer, S., Leithinger, D., Olwal, A., Hogge, A., & Ishii, H. (2013, October). inFORM: dynamic physical affordances and constraints through shape and object actuation. In *Uist* (Vol. 13, pp. 417-426).
8. Gomes, A., Rubens, C., Braley, S., & Vertegaal, R. (2016, May). BitDrones: towards using 3D nanocopter displays as interactive self-levitating programmable matter. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 770-780). ACM.
9. Guinness, D., Cutrell, E., & Morris M.R. (2018, May). Caption Crawler: Enabling Reusable Alternative Text Descriptions Using Reverse Image Search. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM.
10. Guinness, D., Szafir, D., & Kane, S. K. (2017, June). GUI Robots: Using Off-the-Shelf Robots as Tangible Input and Output Devices for Unmodified GUI Applications. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 767-778). ACM.
11. Jeon, M., Winton, R. J., Yim, J.-B., Bruce, C. M., & Walker, B. N. (2012). Aquarium fugue: interactive sonification for children and visually impaired audience in informal learning environments. *Proceedings of the 18th International Conference on Auditory Display (ICAD 2012)*, 246-247.
12. Kane, S. K., & Bigham, J. P. (2014, March). Tracking@ stemxcomet: teaching programming to blind students via 3D printing, crisis management, and twitter. In *Proceedings of the 45th ACM technical symposium on Computer science education* (pp. 247-252). ACM.
13. Kane, S. K., Morris, M. R., Perkins, A. Z., Wigdor, D., Ladner, R. E., & Wobbrock, J. O. (2011, October). Access overlays: improving non-visual access to large touch screens for blind users. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 273-282). ACM.
14. Kestner, J., Leithinger, D., Jung, J., & Petersen, M. (2009, February). Proverbial wallet: tangible interface for financial awareness. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction* (pp. 55-56). ACM.
15. Le Goc, M., Kim, L. H., Parsaei, A., Fekete, J. D., Dragicevic, P., & Follmer, S. (2016, October). Zooids: Building blocks for swarm user interfaces. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 97-109). ACM.
16. MacLeod, H., Bennett, C. L., Morris, M. R., & Cutrell, E. (2017, May). Understanding Blind People's Experiences with Computer-Generated Captions of Social Media Images. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 5988-5999). ACM.
17. Mayer, R.E. & Moreno, R. *Educational Psychology Review* (2002) 14: 87.
<https://doi.org/10.1023/A:1013184611077>
18. Morash, V., Siu, Y.-T., Miele, J.A., Hasty, L., & Landau, S. (2015). Guiding novice web workers in making image descriptions using templates. *ACM Transactions on Accessible Computing*, 7(4), 12:1-12:21.
19. Nowacka, D., Ladha, K., Hammerla, N. Y., Jackson, D., Ladha, C., Rukzio, E., & Olivier, P. (2013, April). Touchbugs: Actuated tangibles on multi-touch tables. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 759-762). ACM.
20. Özgür, A., Lemaignan, S., Johal, W., Beltran, M., Briod, M., Pereyre, L., Mondada, F. and Dillenbourg, P. (2017, March). Cellulo: Versatile Handheld Robots for Education. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 119-127). ACM.

21. Plimmer, B., Crossan, A., Brewster, S. A., & Blagojevic, R. (2008, April). Multimodal collaborative handwriting training for visually-impaired people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 393-402). ACM.
22. Salisbury, E., Kamar, E., and Morris, M.R. Toward Scalable Social Alt Text: Conversational Crowdsourcing as a Tool for Refining Vision-to-Language Technology for the Blind. *Proceedings of HCOMP 2017*.
23. Stangl, A., Kim, J., & Yeh, T. (2014, June). 3D printed tactile picture books for children with visual impairments: a design probe. In *Proceedings of the 2014 conference on Interaction design and children* (pp. 321-324). ACM.
24. Suzuki, R., Stangl, A., Gross, M. D., & Yeh, T. (2017). FluxMarker: Enhancing Tactile Graphics with Dynamic Tactile Markers. *arXiv preprint arXiv:1708.03783*.
25. Taylor, B. T., Dey, A. K., Siewiorek, D. P., & Smailagic, A. (2015, October). TactileMaps. net: a web interface for generating customized 3D-printable tactile maps. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (pp. 427-428). ACM.
26. Tomlinson, B. J., Winters, R. M., Latina, C., Bhat, S., Rane, M., & Walker, B. N. (2017). Solar System Sonification : Exploring Earth and its Neighbors through Sound Georgia Tech Center for Music Technology (GTCMT) 2 , Georgia Institute of Technology , Atlanta , USA, (Icad), 128–134.
27. Völkel, T., Weber, G., & Baumann, U. (2008). Tactile graphics revised: the novel brailledis 9000 pin-matrix device with multitouch input. *Computers Helping People with Special Needs*, 835-842.
28. Walker, B. N., Kim, J., & Pendse, A. (2007, August). Musical soundscapes for an Accessible Aquarium: Bringing Dynamic Exhibits to the Visually Impaired. In *ICMC*.
29. Wilson, G., Carter, T., Subramanian, S., & Brewster, S. A. (2014, April). Perception of ultrasonic haptic feedback on the hand: localisation and apparent motion. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1133-1142). ACM.
30. Wu, S., Wieland, J., Farivar, O., & Schiller, J. (2017, February). Automatic alt-text: Computer-generated image descriptions for blind users on a social network service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1180-1192). ACM.
31. YouDescribe – Audio Descriptions for YouTube videos, 2018. <https://youdescribe.org>