Creating Tutorials with Web-Based Authoring and Heads-Up Capture

The authors built a Google Glass-based head-mounted media capture application for authoring tutorials and studied its use compared to standalone (camera-on-tripod) equipment. Tutorial creators preferred the wearable device, especially when recording activities involving larger objects in nontabletop environments.

Scott Carter, Pernilla Qvarfordt, and Matthew Cooper
FX Palo Alto Laboratory

Ville Mäkelä
University of Tampere, Finland

Tools for capturing and authoring tutorials typically support a limited set of media and thus restrict expressiveness. Past work has shown that relying on a single media type is rarely the best way to convey expository content.[1] For certain tasks, video is particularly helpful (relative to static graphics),[2,3] because some tasks involve gradual progressions that can be difficult to capture in static photos—for example, fluffing egg whites. Other tasks, such as playing musical instruments, can require multimedia feedback.

Video can also help coordinate a set of steps into a cohesive sequence. For example, we could depict the act of kicking a football using a series of static shots—lining up the foot, striking the ball at a particular spot, following through, and so on—but without seeing these individual elements combined in a swift strike, it might be difficult to verify the correctness of the composite end result. Furthermore, using video doesn't preclude integrating static content—many video editing tools support the integration of static photos that can be "played" for some period of time within the video. Semiautomated systems can also help condense expository video into more consumable clips.[4] However, past work has also shown that video is not the best format for all learning tasks.[5] In some cases, the best approach involves combining text with static graphics.[6]

For these reasons, mixing different types of media in the authoring process is critical for creating effective how-to videos (howtos). Capturing content in the medium best suited to convey information is similarly important. Our early need-finding work indicated that many tutorial authors had difficulty recording complicated procedures with standard mounted cameras, suggesting a role for head-mounted capture. This approach stands to benefit tutorial users as well, because past work has suggested that first-person video instruction can improve performance on assembly[7] and learning[8] tasks.

In this article, we explore the use of head-mounted capture for tutorial content creation. We built a set of tools for this, including a head-mounted content capture application. Our how-to authoring toolset lets users tailor the media capture, authoring, and annotation approach to the task being documented. Using these tools, we report on the results of a user study focused on the use of head-mounted devices to compose how-to videos.

Published by the IEEE CS ■ 1536-1268/15/$31.00 © 2015 IEEE

ShowHow: A Tutorial Authoring System

In our previous need-finding work, we interviewed nine tutorial creators about their how-to creation practices and were able to observe two participants as they jointly created a how-to guide.[9] We found the following:

- The capture process often is fundamentally subordinate to performing the documented activity itself. Authors are frustrated when their capture devices obstruct or complicate completing the activity. They are also likely to use familiar tools such as their camera phones and are willing to trade capture quality for increased usability and convenience.
- Authors need to be able to choose a capture device suited to the context of the activity—what is being documented and where—to make media capture as unobtrusive as possible.
- Different tutorials have different capture requirements. Tabletop tutorials tend to focus on the construction of smaller objects and involve fewer camera-angle changes and longer shots. On the other hand, how-tos concerning other tasks (such as automotive or home repair) necessarily require a wider variety of camera angles and shorter-duration shots.
- Authors had difficulty creating a narrative using video exclusively. Many content creators wanted to import and link to external content as well as bookmark and annotate important sections of their video. They wanted to link dynamically to other examples or versions of the object or activity they were documenting.

The last item is consistent with the work of Cristen Torrey and her colleagues, who found that how-to "sharing occurs within and across a collection of communication tools without any centralized control."[10]

Based on these findings, we built a Web-based tool, called ShowHow, with support for authoring and annotating tutorials.[9] However, pilot studies revealed it to be frustrating to use. Ultimately, we found that the interface failed to integrate into the flow of users' workaday tasks, largely because it forced users to interact with media indirectly.

We then redesigned ShowHow to emphasize direct manipulation and remixing of multimedia content. Now, for example, users can drag media directly onto the page and our system will automatically create a well-formatted annotation. Importantly, the system supports representing, integrating, and manipulating a wide variety of media formats, including audio, PDF, photo, video, animated GIF, and text. Users can capture content with ShowHow capture applications or a third-party mobile client of their choosing, and use ShowHow's HTML5-based drag-and-drop authoring tools to assemble captured media into a tutorial.

Head-Mounted Capture

Head-mounted devices offer a potential opportunity to shift authors' attention and effort away from capture devices and toward completing the activity to be documented. Holding a mobile device can limit users' range of motion, but even when users place a device on a tripod, they still must frame the content through a viewfinder that might be difficult to see if the author is away from the tripod. Tripod mounts can also be tedious or difficult to move between and especially during shots.

For these reasons, we decided to build a wearable capture application. Our goal was to build as unobtrusive a tool as possible. As previous need-finding work showed, the wearable system shouldn't force the user into unnatural actions or obscure the subject of the tutorial. At the same time, users should be able to predict what the system is going to record. If users are unsure whether their actions are within the captured field of view, they will try to center the activity within the recording frame. Thus, mounting the capture device to the head near the eyes is ideal.

A head-mounted system could be similar to augmented reality (AR) displays. However, unlike AR systems (such as those described by Steven Feiner and colleagues[11] and by Bui Minh Khuong and colleagues[12]) that primarily support tutorial access, in a wearable capture system, the user needs only straightforward methods to start and stop recording, review their recording, snap photos, and so on. That said, simpler head-mounted capture devices such as GoPro cameras (http://gopro.com) fail to provide enough interactive feedback for users. They might also require users to consult a mobile device during and after recording, which can be cumbersome.

In the end, we chose to use Google Glass because it sits between these two extremes. Its design is relatively unobtrusive; for example, Glass is not a goggle-like wraparound but rather a svelte form factor that

- obscures only a small fraction of the visual field;
- does not require wires that could hinder movement during recording; and
- includes a display for simple feedback, a camera near the eye, and modest interaction methods (such as touch commands).

Our aim was to reduce required interaction during content capture and allow how-to authors to focus on their message rather than on the capture tool. However, the built-in camera application on Glass has several drawbacks. First, there is no preview for taking pictures; this forces users to take photos without knowing how well the content of interest is framed. Second, to record clips longer than 10 seconds, users must make a separate selection. Third, at the time of our research, none of the applications available supported uploading captures automatically—they all required the user to specifically select a photo or video and then "share" it with

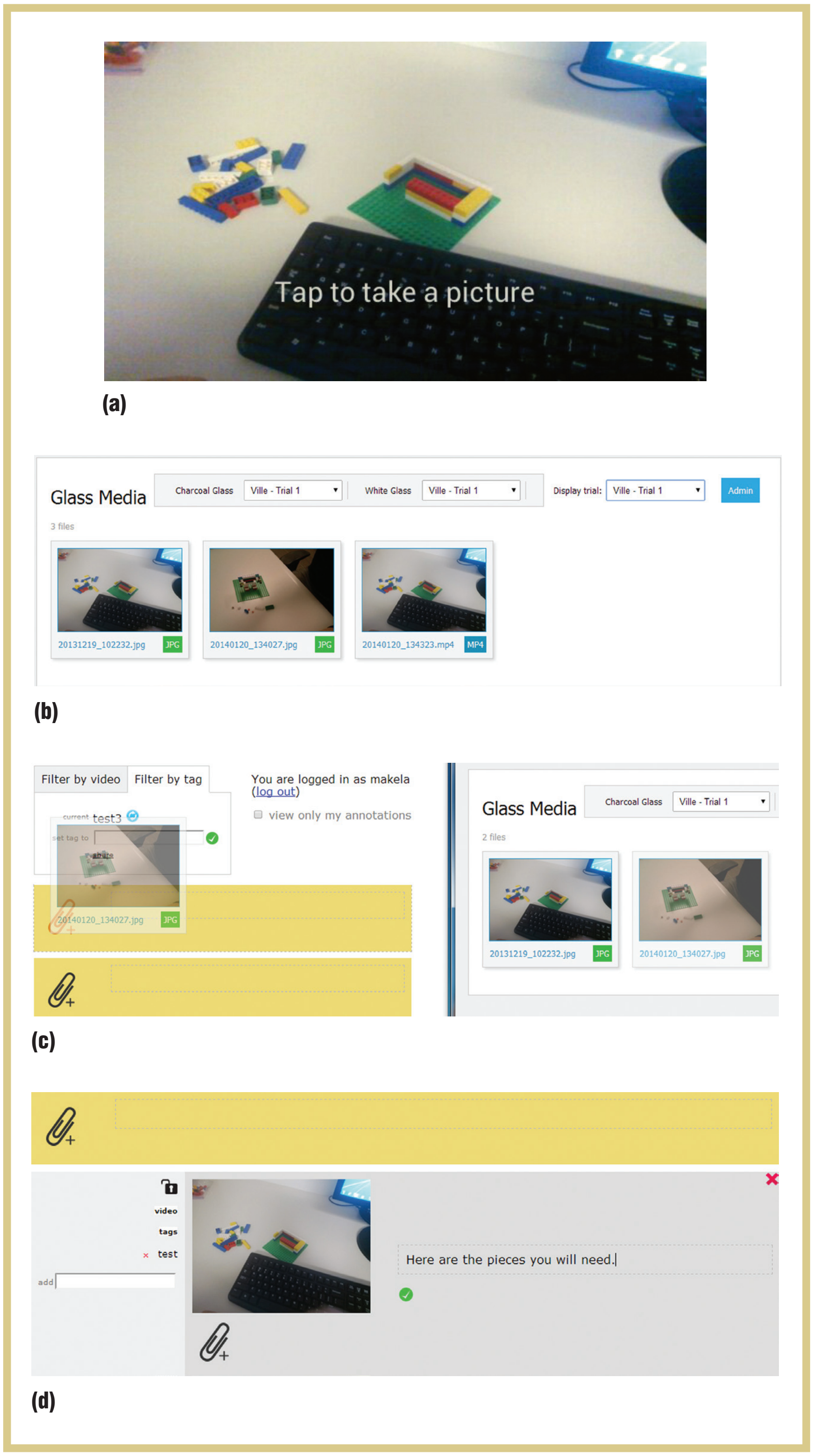


Figure 1. Capturing media with the ShowHow Glass application. (a) Authors snap a photo or record a video using the application. (b) ShowHow automatically synchronizes the media to its webserver and client. (c) From there, authors can drag and drop media into the ShowHow authoring tool. (d) Finally, authors can add text annotations describing the captured content.

an application that would then upload the media to a remote service.

We thus developed our own capture application for Google Glass, which lets users take pictures and record videos with glanceable preview. The application automatically uploads captured content to the ShowHow server and imposes no video duration restrictions (see Figure 1). The ShowHow Glass application is built on an Android platform version 4.0.3. It uses the Glass Development Kit library's gesture detector to recognize different gestures on the Glass touchpad. A separate component handles uploads to the server.

The ShowHow Glass application consists of three views. From the main menu, users select whether they want to take pictures or record videos. To navigate, the user swipes forward or backward on the touchpad attached to the Glass frame and taps the touchpad to make a selection. Users return to the main menu by swiping down on the touchpad. In picture view a single tap takes a picture, while in video view the first tap starts and the second tap stops the recording (Figure 1a). Glass's display shows a viewfinder while in capture mode, making the recording glanceable. Media is uploaded automatically, and the upload progress is updated in the top-right corner of the Google Glass screen. The Glass application uploads files asynchronously, and they appear in the Web application as they become available. After the media files are uploaded, they are visible in ShowHow (Figure 1b) in the order they became available. Then the creator can drag and drop media into the ShowHow authoring tool (Figure 1c and 1d).

Authoring How-Tos with ShowHow

ShowHow supports both video-based and document-based methods for authoring. How-to content is commonly structured according to the steps that comprise the documented activity. Both authoring approaches let users combine different types of media to craft tutorials that optionally include step

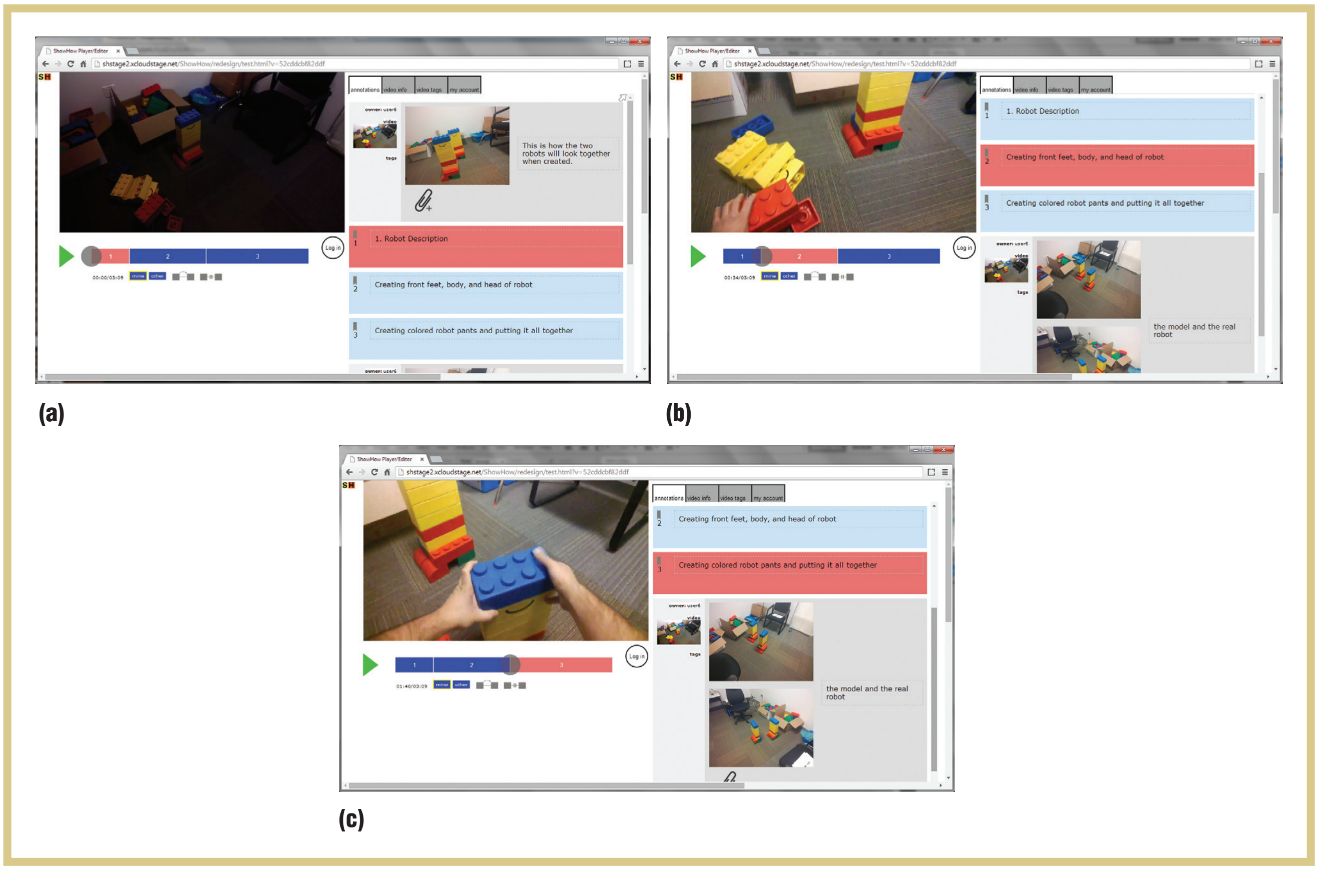


Figure 2. A video-based tutorial created by a participant in the study. This tutorial includes three steps (bookmarks) as well as annotations at the beginning and end with photos showing the full figure (in this case, a completed Lego robot).

boundaries, step sequence, step importance, and references to other content or activity beyond the immediate scope of the current how-to. The key distinction is that the video-based approach fundamentally relies on a temporal organizational metaphor wherein the system plays through the video and associated bookmarks, while the document-based approach relies on a spatial organizational metaphor wherein the user scrolls through content.

The process of creating a video-based how-to involves uploading raw content, editing that content, and then adding bookmarks and multimedia annotations to augment the how-to. To support rapidly combining and remixing short clips, we built an HTML5-based video creator. Authors can drag and drop video clips onto the tool, drag to reorder clips, and play each clip individually. They can also produce a higher-quality clip with any external editing tool and upload it to our system to take advantage of our tools' rich multimedia annotation abilities. Figure 2 shows an example of a video-based tutorial created during our study. When the author is satisfied with the order and edits of individual clips, she clicks a button to create a composite clip. Once processing is complete, she can view the video and begin adding markups.

The document-based approach lays out each step in the process separately using dedicated text and media. Figure 3 shows a document-based tutorial created with ShowHow. To start, a user drags media to the tool. This creates a "tag" where the user can add annotations. As in the video-based case, annotations include text and an arbitrary number of media elements. Each annotation is automatically assigned the current tag as it is created. The temporal sequencing of the video clips within the video-based tutorial corresponds to the vertical arrangement of the multimedia annotations in the document-based tutorial.

The document approach has some advantages over the video approach: steps can be skimmed with a glance, searched more directly, and never need to be "paused." In contrast, the video approach is more dynamic and doesn't require manually advancing step by step. We examine head-mounted capture in conjunction with authoring both types of how-tos.

User Study

To assess the prospects of wearable capture for tutorial creation, we compared our head-mounted capture

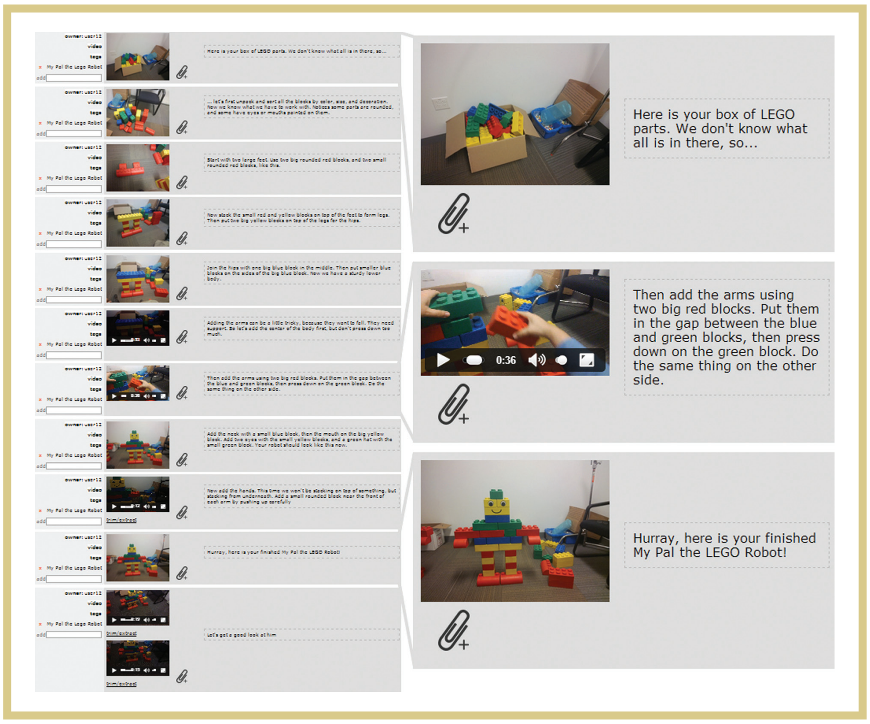


Figure 3. A document-based tutorial created by a participant in the study. This 11-step document uses a combination of photos, videos, and rich text to illustrate the construction of a robot using large blocks. The first, seventh, and tenth steps are shown in detail (on the right). Note that in the seventh step, the video starts at 36 seconds, indicating that the tutorial creator trimmed the original video clip.

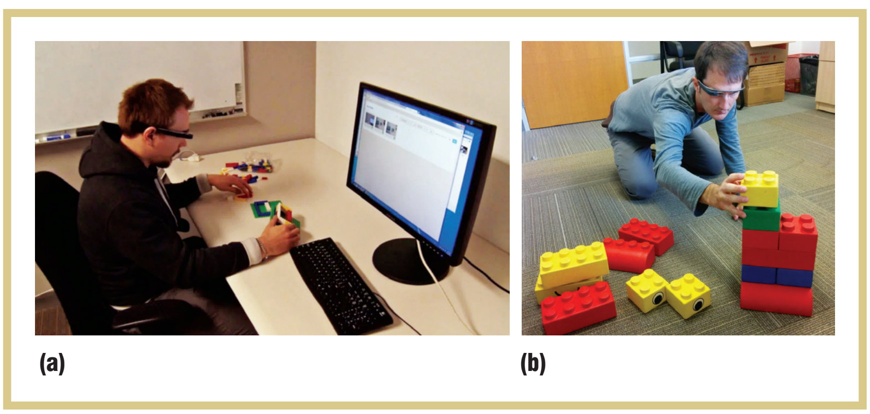


Figure 4. Creating tutorials that demonstrate how to build a robot with Lego blocks. (a) A user completing the tabletop task while recording with Glass. The media is synchronized automatically to the webpage shown on the monitor. (b) For the floor task, participants used foam blocks much larger than their standard Lego counterparts.

system with portable cameras on tripods in two common tutorial creation scenarios: working at a table and freely moving around a larger object. These two situations impose different requirements that influence how creators approach and experience the media capture process. A camera on a tripod can require extensive manipulation, which might limit the number of angles a person can choose for recording objects of interest; however, it can produce higher-quality media. Our study focuses on users' experiences with head-mounted capture and tripod-based capture in both of these settings.

As capture devices, we used Google Glass running the ShowHow Glass application and a Nexus 4 camera on tripod. We set up the Nexus 4 to automatically save media to a Dropbox folder, which would upload the media to a computer for editing, simulating the automatic upload implemented in our ShowHow Glass capture application. In both conditions, participants authored a tutorial by dragging media into the ShowHow Web authoring tool from either the ShowHow Web client (Glass) or the Dropbox folder (Nexus 4).

We asked 12 people to each create two tutorials (in separate sessions) demonstrating how to build a robot at least eight layers tall with Lego blocks. They were free to choose the structure and focus of their tutorials, but the final versions of the tutorials had to enable viewers to build an exact replica. In one tutorial, the participants sat at a table and used regular-size Lego blocks (see Figure 4a); in the other, they used larger foam bricks on the floor (see Figure 4b). The participants used different capture devices for the two tutorials: Google Glass and a camera on a tripod. During the design phase, we encouraged the participants to try out the recording device and authoring tool. They were also asked to frame the videos and photos to best suit their needs by moving the tripod or changing the recording position while wearing Google Glass. After each tutorial, participants answered questionnaires about their experience with our tools. The participants spent on average 23 minutes, with a standard deviation (SD) of 15.0 minutes, per study session.

We created our authoring tools for amateur tutorial producers. Since part of our aim was to explore the ease of

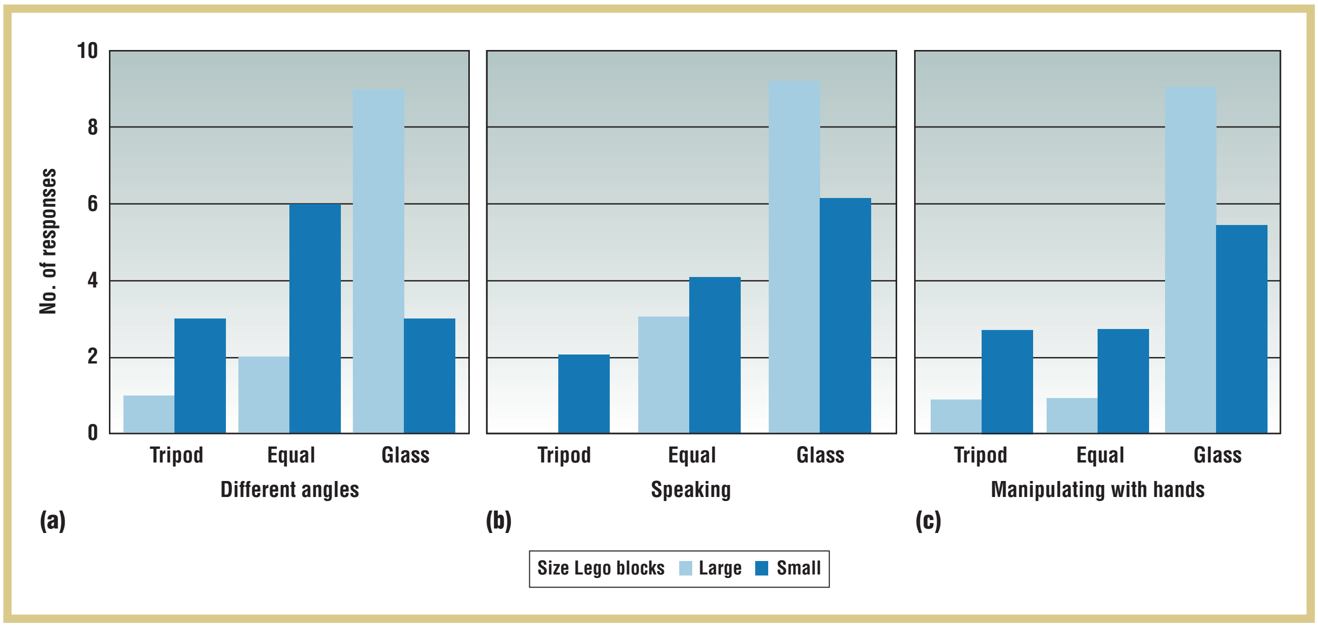


Figure 5. The number of participants preferring the camera on tripod, Google Glass, or equal (no) preference when (a) showing their small and large models from different angles, (b) showing their small and large models and speaking (nobody preferred using the tripod to add speech to the video while working with large Lego blocks), and (c) showing their small and large models and manipulating an item with their hands.

use of the toolset, we used infrequent tutorial producers for our study. All participants had experience with video editing and making tutorials, although most (10) participants made tutorials less than once a year. The remaining two made up to six tutorials per year.

The order and combination of the three independent variables (activity location, capture device, and authoring method) was balanced over all participants. All participants used both capture devices and both authoring methods (video-based and document-based).

Results and Discussion

Many factors influence the choice of capture devices and authoring tools. In our discussion, we focus on the creators' media capture experience and the quality and characteristics of the captured media.

Experience with the Capture Devices

We asked the study participants to compare the capture devices in activities with small versus large models (see Figure 5). We found that participants significantly preferred Glass for capturing a large model but reported equal preference for the two devices when capturing a small model (using the Wilcoxon paired signed rank test, $Z = -2.16$, $p < 0.05$). They also preferred head-mounted capture while speaking and while working with their hands. Participants' comments reflected feeling comfortable with head-mounted capture when working with large objects or actively moving. Five of the participants felt that the camera and tripod would be best suited for recording stable video and when no movement or different angles are required, such as detailed work in a static location.

For content creators, feeling confidence in the capture method is critical. Our data indicates that participants found head-mounted capture straightforward to use, provided good feedback about what was in view, and required minimal setup. Specifically, half of the participants said they appreciated that the camera would share their field of view while building the robots. The following are representative statements from two participants:

- "Once I got the hang of it, I was able to check with a glance the recorded video framing/quality etc. without taking my eyes off the model/task at hand" (Participant 2).
- "It's more natural and what you see is what you get so less time [is] wasted in adjusting the camera or the object" (Participant 4).

Users also reported some difficulties using Glass to capture photos and videos. Three participants felt they had to regularly check that Glass's camera was recording what they wanted. One participant said, "Occasionally this distracted from the task as I was looking at what the viewfinder was displaying" (Participant 8). Those who didn't check the viewfinder during recording sometimes noticed later that the objects they intended to show in the video were outside the frame, either partially or completely. Although Glass's line of sight matches well with the wearer's, five participants pointed out that this is not always sufficient; Participant 10 said, "Sometimes I forgot to keep my head pointed to my hands." Also participants noted that when recording smaller objects, Glass had no ability to zoom in (noted by two of the 12 participants—that is, 2/12) or show certain angles (2/12). Moving the head closer to the object might not feel natural or comfortable. Participant 4 explained, "Since you're wearing Glass, certain angle e.g. near the floor is hard to record." We noticed one participant tilting his head to get a good close view when trying to capture a Lego block close to the floor. However, this did not result in a good shot of the block when the video image suddenly tilted.

Participants had mixed feelings about the camera-on-tripod setup. Four participants considered this setup convenient, while others focused on drawbacks such as the difficulty seeing what the camera was recording (5/12), the need for additional adjustment of the camera angle (4/12), and the constraint of working within the camera's field of view (3/12). Here are two representative comments:

- "I wasn't sure if the object was in the field of view of the camera, and it was kind of complicated to check if the object is being viewed and document the build process at the same time" (Participant 8).
- "The initial tripod setup was non-trivial and I was less inclined to mess with it once it was set up and so I ended [up] moving the model around and closer to the camera to get different angles and better shots" (Participant 2).

According to our participants, the main advantages of using a camera on a tripod were good control of the camera angle (5/12) and the good quality of the resulting video (4/12).

Our results also indicate that tutorial creators value the convenience and confidence Glass gives them over potential quality issues. This tradeoff was particularly evident when working with a larger model that required movement over a larger area. Setup with Glass is minimal: the creator can simply move to a new position. Feedback of how well the content is framed is accessible with a glance. Video can become more shaky than when capturing with a camera on tripod, but the value of unrestricted movements and instant feedback was clearly higher for our participants.

Captured Media

To explore how participants used the tools available to them, we examined the media captured for the tutorials, and how participants manipulated the camera or robot to show different perspectives. On average, participants captured 8.5 photos (SD = 10.80) with Glass and 7.0 photos (SD = 5.84) with a camera on tripod, and 3.1 video snippets (SD = 3.68) with Glass and 4.2 video snippets (SD = 2.86) with a camera on tripod. The total video duration was 144 seconds (SD = 79.7) for video captured with Glass and 148 seconds (SD = 143.5) for video captured by a camera on tripod. None of these differences were significant. The location did not impact either the number of photos or video snippets or the captured video duration. The tutorial's topic dictates to some extent how much media the participants need to capture to convey their intended message. In this study, we held the topic constant so that we could examine whether different capture tools produce different strategies for capturing the necessary media.

An object can be depicted from different angles by either moving the camera or moving the object itself. When working with regular-sized (smaller) Lego blocks, the participants rarely moved the camera and tripod (M = 0.4 times, SD = 0.79) and instead manipulated the Lego model to show multiple views. The participants were just as likely to move the Lego model when capturing with Glass as with a camera on tripod, 4.6 moves per tutorial (SD = 6.29). Moving a large object is more difficult, and thus the participants moved the camera and tripod—on average 4.6 times (SD = 5.53) with the larger bricks. Clearly, the need to frame the Lego robot from different perspectives was equal when creating the tutorial on the tabletop and on the floor; only the method differed.

When focusing on shifts in framing (such as changes in angle, pan, or zoom) within the recorded videos, we found that participants made significantly more framing shifts per minute (M = 6.0 shifts, SD = 4.78) using Glass compared to using a camera on tripod (M = 1.4, SD = 2.91; $F(1, 10) = 15.00$, $p < 0.01$). In fact, 69 percent (SD = 34.7) of the framing shifts in videos captured by Glass were made within a video snippet. When capturing with a camera on tripod, the participants made 33 percent (SD = 37.5) of the framing shifts within a video snippet. In addition, the participants made most of these framing shifts while working with the large Lego blocks on the floor (M = 6.2 shifts per minute, SD = 5.02), compared to when they were using the tabletop (M = 1.2 shifts per minute, SD = 1.76, $F(1, 10) = 18.2$, $p < 0.01$). The large majority of the framing shifts seated at the table were done when the participants were using Glass (M = 2.3, SD = 1.91). While capturing with Glass, the framing shifts were sometimes an unintentional byproduct of participants' head movements.

On average, the participants retained 89 percent of the captured video in their final edited tutorials. The participants often trimmed the start and end point of a video snippet, as these generally contained more inadvertent framing shifts. Overall, these results show that participants using Glass to record larger objects more often changed the framing of the object to better convey important aspects of the building process compared to when they used a camera on tripod. The reason might be that a change in framing can easily be done by moving the head or the body when wearing Glass. The threshold for doing a framing shift with a camera on tripod appears to be considerably higher.

Tutorial Media Quality

We asked three judges, professional and amateur photographers and videographers, to evaluate each tutorial's media quality. They rated the media used in the tutorials (photo, video, and audio) on a seven-point scale based on sharpness, focus, and framing. Media quality is of course only one aspect of the tutorial's overall effectiveness. However, it is important because the media must illustrate important steps. If the video is too shaky or badly framed, or the audio is inaudible, a tutorial consumer will have more trouble following it. Of the 10 top-rated tutorials on media quality, seven were recorded with Glass and three with a camera on tripod. In comparison, of the 10 tutorials with the lowest ratings, four were recorded with Glass and six were recorded with a camera on tripod.

These results are interesting because the participants perceived improved recording quality when using the camera on a tripod. A head-mounted camera is more sensitive to head movements, and the camera angle on Google Glass is not quite designed for recording manipulations with the hands close to the body. For this kind of recording, the participants needed to consciously frame the capture. However, through the complete authoring process, the participants managed to create tutorials with equal or better quality using head-mounted capture than using a camera on tripod.

In sum, users found head-mounted capture at least as effective as using a camera on tripod across our study conditions, and they preferred it for larger activities. They also preferred head-mounted capture for adding natural narration from multiple vantage points. Working at a tabletop, using a camera on a tripod is an equally good alternative. The content they obtained using head-mounted capture in turn was judged to be of equal or higher quality than that of the camera on a tripod.

Creating a tutorial requires capturing descriptive and procedural content and rendering it so that viewers can comprehend and recreate the steps taken. In our study, we have seen that seamless capture of tutorial content lets the authors focus on exposition. Head-mounted capture was perceived as easy to use, having a minimal setup with good feedback, and a first-person view of the content. The Glass application was seen as particularly useful for creating tutorial content involving large objects where authors needed to move around to show different viewpoints.

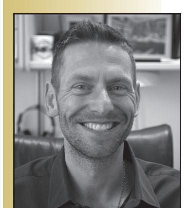
Our findings support the idea that head-mounted devices that integrate capture and some form of real-time feedback "could introduce a new form of interactive manual into the world."[13] Once capture tools are paired with complementary access applications (which we are in the process of building), users will be able to view and navigate both video- and document-based tutorials on a head-mounted display and capture system. With this end-to-end system, we might come to use these devices to construct threaded conversations with family, friends, and colleagues about how to interact with everyday objects in our world.

Additionally, we found that a close integration between capture and composition tools streamlines the authoring process. Our suite of Web tools enables users to combine a variety of media to convey tutorial material. We

the AUTHORS

Scott Carter is a senior research scientist at FX Palo Alto Laboratory. His primary research focus is developing innovative multimedia user interfaces. He received his PhD in computer science from UC Berkeley. Contact him at carter@fxpal.com.

Pernilla Qvarfordt is a senior research scientist at FX Palo Alto Laboratory. Her research focuses on understanding human attention and behavior when interacting with multimedia environments. She received her PhD in computer science from Linköping University, Sweden. Contact her at pernilla@fxpal.com.



Matthew Cooper is a senior research scientist at FX Palo Alto Laboratory where he leads the Social and Enterprise Media group. He develops content analysis techniques to support interactive multimedia applications. He received his DSc in electrical engineering from Washington University in St. Louis. Contact him at cooper@fxpal.com.

Ville Mäkelä is PhD student in the School of Information Sciences at the University of Tampere in Finland. His research interests include public displays, gestural interaction, ubiquitous environments, and augmented reality. He received his M.Sc. in interactive technology from the University of Tampere. Contact him at ville.mi.makela@sis.uta.fi.

intend to evaluate how the different metaphors in the authoring model impact how and what tutorial creators capture. We believe it's important to understand the whole cycle of capture, authoring, and tutorial access to be able to design tools suited for the unique multimedia capture and consumption requirements of tutorial content.

REFERENCES

1. S. Feiner and K.R. McKeown, "Automating the Generation of Coordinated Multimedia Explanations," *Computer*, vol. 24, no. 10, 1991, pp. 33–41.
2. T. Grossman and G. Fitzmaurice, "ToolClips: An Investigation of Contextual Video Assistance for Functionality Understanding," *Proc. ACM Conf. Human Factors in Computing Systems* (CHI), 2010, pp. 1515–1524.
3. S. Pongnumkul et al., "Pause-and-Play: Automatically Linking Screencast Video Tutorials with Applications," *Proc. 24th ACM Symp. User Interface Software and Technology* (UIST), 2011, pp. 135–144.
4. P.-Y. Chi et al., "DemoCut: Generating Concise Instructional Videos for Physical Demonstrations," *Proc. 26th ACM Symp. User Interface Software and Technology* (UIST), 2013, pp. 141–150.
5. S. Palmiter, J. Elkerton, and P. Baggett, "Animated Demonstrations vs. Written Instructions for Learning Procedural Tasks: A Preliminary Investigation," *Int'l J. Man-Machine Studies*, vol. 34, no. 5, 1991, pp. 687–701.
6. S.M. Harrison, "A Comparison of Still, Animated, or Nonillustrated Online Help with Written or Spoken Instructions in a Graphical User Interface," *Proc. ACM Conf. Human Factors in Computing Systems* (CHI), 1995, pp. 82–89.
7. R.E. Kraut, S.R. Fussell, and J. Siegel, "Visual Information as a Conversational Resource in Collaborative Physical Tasks," *Human-Computer Interaction*, vol. 18, no. 1, 2003, pp. 13–49.
8. R. Lindgren, "Generating a Learning Stance through Perspective-Taking in a Virtual Environment," *Computers in Human Behavior*, vol. 28, no. 4, 2012, pp. 1130–1139.
9. S. Carter et al., "Tools to Support Expository Video Capture and Access," *Education and Information Technologies J.*, vol. 19, no. 3, 2013, pp. 637–654.
10. C. Torrey et al., "HowTo Pages: Informal Systems of Expertise Sharing," *Proc. European Conf. Computer-Supported Cooperative Work* (ECSCW), 2007, pp. 391–410.
11. S. Feiner, B. MacIntyre, and D. Seligmann, "Knowledge-Based Augmented Reality," *Comm. ACM*, vol. 36, no. 7, 1993, pp. 53–62.
12. B.M. Khuong et al., "The Effectiveness of an AR-Based Context-Aware Assembly Support System in Object Assembly," *Proc. IEEE Virtual Reality* (VR), 2014, pp. 57–62.
13. M. Svenvold, "The Disappearance of the Instruction Manual," *Popular Science*, Feb. 2015; www.popsci.com/instructions-not-included.