Polly: Telepresence from a Guide's Shoulder

Don Kimber[1(✉)], Patrick Proppe[1], Sven Kratz[1], Jim Vaughan[1], Bee Liew[1], Don Severns[1,2], and Weiqing Su[2]

[1] FX Palo Alto Lab, California, USA
donkimber@gmail.com
[2] University of California, San Diego, San Diego, USA

Abstract. Polly is an inexpensive, portable telepresence device based on the metaphor of a parrot riding a guide's shoulder and acting as proxy for remote participants. Although remote users may be anyone with a desire for 'tele-visits', we focus on limited mobility users. We present a series of prototypes and field tests that informed design iterations. Our current implementations utilize a smartphone on a stabilized, remotely controlled gimbal that can be hand held, placed on perches or carried by wearable frame. We describe findings from trials at campus, museum and faire tours with remote users, including quadriplegics. We found guides were more comfortable using Polly than a phone and that Polly was accepted by other people. Remote participants appreciated stabilized video and having control of the camera. One challenge is negotiation of movement and view control. Our tests suggest Polly is an effective alternative to telepresence robots, phones or fixed cameras.

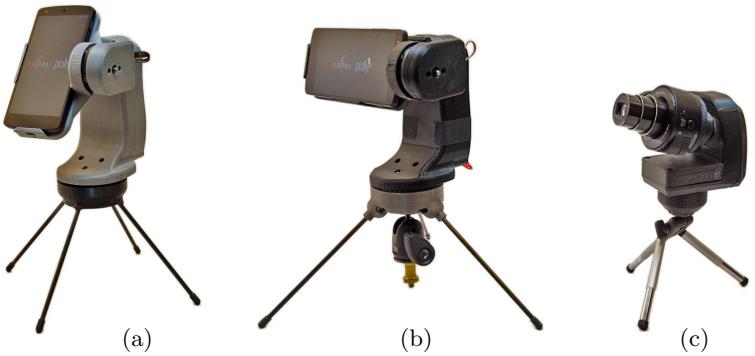
Keywords: Telepresence · Image stabilization · Remote guiding · Wearable · Gimbal · User feedback · Iterative design

1 Introduction

According to Ryan and Deci's [1] research, human's intrinsic motivation to thrive is based on psychological needs. Sheldon et al. list ten basic psychological needs, including two "most fundamental" needs of *pleasure-stimulation* and *relatedness* [2]. Pleasure-stimulation is addressed by finding sources for new experiences of sensations and activities which cause pleasure and enjoyment. Relatedness is the need to be close to people who are important to yourself and loved ones. This might be one psychological reason why most people have a desire to go outside and visit interesting locations (e.g. museums, zoos or parks) and also to experience events such as family gatherings or music concerts together with other people. However, not everyone is able to fulfill their wishes to explore or be able to be close to their kin. Reasons for this may be that these persons are physically immobile, sick in a hospital or simply very far away from where they would like to be.

To address this issue of mobility and presence, we have developed a wearable system called *Polly*, motivated by the metaphor of a remotely controlled parrot which could rest on someone's shoulder and look around independently of the

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part III, LNCS 8927, pp. 509–523, 2015.
DOI: 10.1007/978-3-319-16199-0_36



(a) (b) (c)

Fig. 1. From left to right: (a) Portrait Polly, (b) Landscape Polly, (c) Sony LensCam Polly. All prototypes are based on a three-axis brush-less gimbal that can be carried, placed on perches or flat surfaces, or worn by the guide at shoulder level.

view of the person carrying it. [1] We have developed a series of Polly prototypes and our current version consists of a three-axis stabilization gimbal driven by brush-less motors holding a mobile phone that provides the audio and video feed and a connection to the internet. Polly can be carried by hand, can be placed on 'perches', can rest on surfaces, or be worn by way of a backpack frame with an attachment holding Polly near the shoulder of its wearer. We implemented applications that let remote users control the gimbal orientation (pitch and yaw) from their location. We used Skype, Vidyo and Google Hangouts for audio and video transmission, but any suitable software with a mobile client suffice. [2]

We chose a wearable solution since mobile robots are difficult to control and lack mobility over terrain that is not adapted to their style of locomotion (e.g. staircases). The solution we believe that works best with the current level of technology is to use a human 'guide' at the location a remote person wishes to visit.[3] Having a human guide who carries or wears the Polly device has the following advantages: (1) the guide is in constant communication with the remote user and can easily understand their wishes, (2) humans are extremely mobile and agile, especially in environments built by and for humans, (3) the guide can mediate conversations between the remote operator and other people encountered and (4) the social interaction between the guide and remote person may be a positive part of the overall experience.

The phone is stabilized by a gimbal and can be worn on the shoulder for the following reasons. Firstly, we thought it would be advantageous for the guide to have two free hands and, secondly, not worry about pointing the camera, but allow the remote operator to do this. Finally, we found that if the mobile

[1] Also motivated by the remarkably stabilized head pose of many birds. http://www.youtube.com/watch?v=UytSNlHw8J8

[2] http://www.skype.com, http://www.vidyo.com, http://www.google.com/hangouts/

[3] The term *guide* is for expositional brevity. In some scenarios, the remote person may be more familiar with and knowledgeable about the space being explored.

phone is worn on a lanyard or mounted rigidly to the wearer, the video quality is significantly reduced while the guide is walking, due to excessive movement in the image. The brush-less gimbal uses an IMU for active rotation compensation and leveling resulting in a very smooth video feed, much like a Steadicam rig, but much smaller in size and weight.

After a review of related work in Section 2, Section 3 describes the design process of Polly from initial prototype to its current form, followed in Section 4 by the evolution of the User Interface. Section 5 describes field tests, including five tests with Henry Evans, a quadriplegic experienced with assistive and telepresence robotics who used and helped critique our system and one test with another quadriplegic familiar with telepresence technology. Although Sections 3-5 are presented sequentially they describe aspects of a cyclical interactive design process. In Section 6 we contribute design guidelines for builders of such systems as well as a set of practical recommendations based on our experience, followed by a discussion of further work, and finally our conclusions.

2 Related Work

Drugge et al. proposed a wearable telepresence system consisting of an HMD (Head Mounted Display) and a head-mounted camera [3]. In contrast to Polly, this does not afford the remote participant the same degree of control, since view direction cannot be changed, as it is the case using Polly's gimbal. Mayol et al. describe a wearable camera system and discuss benefits of decoupling the camera orientation from body pose of the wearer, but the control paradigm is based on active vision rather than remote control [4,5]. Similarly, systems such as Google Glass streaming to a Hangouts, or the Tele-actor system [6] do not give the remote user any direct control over view. The MH2 [7] is a shoulder-worn humanoid telepresence robot with a focus on conveying gestures and poses made by the remote remote participant. Polly, on the other hand, is not based on physical representation. Rather, a camera image of the remote participant, displayed via a smartphone, is used to represent him or her. TEROOS [8] is a shoulder-mounted wearable telepresence system, that is perhaps the wearable telepresence system with the closest resemblance to Polly, as it uses servo motors that are controllable by the remote operator. There are, however, two main differences: Firstly, the platform of TEROOS camera does not appear to be stabilized, which, as we found out during initial tests of Polly, produces low quality video while the local operator is walking. Secondly, TEROOS follows the avatar concept, similar to MH2 and uses an abstracted form of representation involving the shape of a decorated camera. Again, we believe that showing the remote operator directly may be advantageous, for example where the mobile user knows the person he is interacting with personally. The MeBot [9] is a small expressive robotic device that does include a display, but is not intended as a wearable device.

There are many telepresence robot systems (see [10] for a survey) and it is becoming more common to use these for providing access to the disabled [11].

(a) (b) (c) (d)

Fig. 2. Polly Prototypes: (a) first frame mounted version, (b) first stabilized version, (c) portrait phone, (d) current landscape version

Some museums are starting to use these systems for telepresence tours, available to limited mobility visitors [12]. These systems are typically many thousands or tens of thousands of dollars and not well suited for outside tours, steps, crowded spaces, etc. By contrast, the cost for a Polly-type device is a few hundreds of dollars and being human carried, they are suitable for outside use. Another project with similar goals to Polly is Virtual Photowalks [13] which matches up photographers able to provide photo walks through beautiful or interesting locations, to remote participants, particularly but not exclusively those constrained by physical disabilities. The remote participants may talk with the photographer and request photos to be taken.

3 Design Evolution of Polly Prototypes

A design goal for Polly was to produce relatively inexpensive devices, costing several hundreds of dollars, to experiment with scenarios allowing one person to provide a video view for a remote person. The baseline case for these scenarios is what people currently do when no specialized solution is available - they run a videoconferencing app on a phone, tablet or laptop and walk around carrying these devices while trying to communicate with their remote friends. Our baseline experience consisted of using a phone in this manner, where the phone was hand held, or held by a simple strap. The videoconferencing apps we tried were Skype, Vidyo and Google Hangouts. Although none of the apps clearly dominated the others in all respects, overall we had better success in terms of video quality and persistence of connections using Skype. On the other hand we found Google hangouts slightly more convenient in terms of supporting new users, providing an integrated remote control interface and supporting multiple video clients.

3.1 First Frame Mounted Version

The first body-mounted prototype (Figure 2(a)) consisted of a fixed shoulder mount frame with no stabilization gimbal and a small daypack that could carry

a tablet, digitization hardware and a battery. We created two versions, one using a Logitech 920C webcam, connected to a Microsoft surface running Skype. The other version used a GoPro camera which could record HD video onto a SD card, while simultaneously outputting analog video, which was digitized using a USB video capture device. We tried several vendors, including Diamond and Hauppauge but none of them worked directly as a video input for Skype. However, we found that using third party virtual camera programs such as XSplit or ManyCam [4] enabled us to use video from the GoPro as an input to Skype.

3.2 First Stabilized Version

We found two main drawbacks to the baseline and first mounted version. One was the lack of camera direction control by the remote participant. The second was that during walking the video was very shaky and unpleasant to watch. To address these, the next version (Figure 2(b)) used a gimbal which provides a sort of Steadicam type stabilization and for which the camera can be pointed remotely. The gimbal used was a three axis brush-less motor gimbal, designed for a GoPro camera on a small UAV (Unmanned aerial vehicle) and using the 8bit version of the SimpleBGC[5] board. The camera direction control inputs to the gimbal are provided as PWM (Pulse Width Modulation) signals, which output from a Pololu USB to PWM device[6].

As with our first frame mounted prototype, the stabilized GoPro version requires a separate device for running Skype and because the GoPro output was analog, the video needed to be digitized. (GoPro also outputs HDMI, but we could not find a portable solution for making this available as Skype input.) One limitation is that wires carrying video signals from the cameras must cross three stabilization motor axes, so it is not possible to get full travel for the motors. Furthermore, the PWM signals from the Pololu device needed to cross the yaw axis. We considered a modified version to address this problem with slip rings, but found problems with high frequency noise caused by the slip rings.

3.3 Second Stabilized Version - Portrait Mode Smartphone Polly

The main deficiency of the first stabilized version was the complexity of the system requiring the camera, gimbal, multiple USB devices and MS Surface tablet, in addition to workaround programs such as XSplit that were necessary together with basic videoconferencing programs. We decided a simpler more usable device would be an entirely self-contained unit consisting only of the gimbal and a smartphone in portrait orientation (Figures 1 and 2(c)). In this design, the camera, videoconferencing software and all necessary control software runs on the phone. The 3D-printed gimbal case contains a battery, a SimpleBGC board, and a Bluetooth 4.0 BLE-PWM device, allowing a Polly control app running on

[4] http://www.xsplit.com/, http://www.manycam.com/
[5] http://www.basecamelectronics.com/SimpleBGC/
[6] A Pololu Micro Maestro 6-Channel USB Servo Controller was used.

the phone to receive messages sent from remote users and control the gimbal via bluetooth. We also included an external bluetooth loudspeaker/microphone which can be worn or mounted to the carrying rig. Another advantage of the self-contained unit is modularity. Polly can be easily carried by snapping it onto a shoulder mounted rig, but can also be carried by hand, or can rest on its own 'feet', all while remaining fully functional and being remotely controlled.

Once this second Portrait Polly iteration was implemented, we started using it in field tests with outside people. We discovered two major issues with this Polly version. First, the maximum range of motion (240 degrees for yaw and 75 degrees for pitch) limited the users feeling of control (e.g. one comment was "I wanted to look down into the ravine.") Second, the phone's camera is in portrait orientation, resulting in video with smaller width than height, in contrast to common cinematic aspect ratios where the with is always greater than the height. This narrow horizontal field of view 'wastes' a lot of pixels on height and does not mimic peripheral vision.

3.4 Current version - landscape mode smartphone Polly

To address the issues of the small range of motion and narrow field of view, we incorporated a landscape mounted phone and different approach to send control commands to the SimpleBGC board into the next and current Polly iteration.

Redesigning a landscape mode gimbal was straightforward due to the modularity and 3D-printed parts, and provided greater pitch axis travel, because the vertical backing of the gimbal case restricts the phone from extreme pitch values in portrait mode, but not in landscape mode. To utilize the increased range of motion, we adjusted the SimpleBGC settings, replaced the BLE-PWM device with a Bluetooth serial module, and implemented simpleBGC's serial command protocol into our Polly control app. The serial protocol now allows a bidirectional communication path and therefore we are able to access the gimbal IMU values and board settings on the android phone. The range of motion was increased up to full 360 degrees in yaw and 180 degrees in pitch, depending on the mode used (see Section 4.2). We have also incorporated a multipurpose mount with an adjustable ball joint, and are in the final stages of implementing a special charging 'perch' which can provide power, allowing permanent operation, instead of a 90 minute battery limitation.

In our field tests, we found that in noisy or windy environments, the external bluetooth loudspeaker and microphone combination is not powerful enough for those settings, and have replaced them with a higher powered speaker.

3.5 Experimental Version

During several experiments with Polly, users expressed the desire to view distant objects. The devices used in these experiments did not have optical zoom capabilities. Video streaming quality was also often poor, regardless of which videoconferencing software we used. We found no solution capable of streaming clear and real-time low latency HD video (via 3G/4G). Our experimental Polly

approach is based on the Sony DSC-QX10, which is a smartphone attachable lens-style camera with up to 10x zoom. This device is capable of live-streaming low latency video over WiFi, while simultaneously recording video in Full HD resolution onto a SD card. Newer Sony cameras, like the DSC-QX10, can be accessed and controlled via a dedicated API. Unfortunately, there is no to make the live-streamed images from the Sony camera available as video input to Skype or Hangouts apps on the phone, particularly without rooting the phone. Furthermore, the camera can only stream images on its own dedicated WiFi, and the android phone can not simultaneously access this WiFi and maintain internet connectivity. To address these limitations, we added a Linux laptop to our setup, running a kernel module to create a V4L2 loopback device [7] accessible as a 'virtual web-camera' to make the Sony camera video stream available as input to Skype or Hangouts on Linux. The laptop uses two wireless adapters to access the camera and internet simultaneously.

We have built this version of Polly, but not yet begun field tests. Although it requires a laptop or other Linux device, it has several attractive features beyond optical zoom and recording capability. It is much smaller in size and weight (0.5kg *vs* 1.0kg, see Figure1(c)) and gives us the capability of running computer vision algorithms such as tracking or object recognition on the video stream.

4 User Interface, Use Modes and Participant Interaction

We describe our approaches and choices for designing the user interfaces, the modes of operation and the way to communicate and negotiate between guide and remote participant, in terms of directions and route planning. This is still one of the biggest issues we are facing while using our prototypes in field tests.

4.1 User Interfaces

One control interface design challenge for Polly is latency, both of control signals and video. This impacts the design choice between 'proportional control' in which a slider or mouse position directly specify camera angle, and 'differential control' in which angular velocity is controlled. We found that for low latency both methods feel natural, but for higher or sporadic latency, differential control is much more difficult to use. In the lab with Polly, servers and control machines all on a single LAN we could assure latency of under .1sec., but over the internet, latencies were larger and sporadic, and were sometimes over a second. So our subsequent interfaces were based on proportional control.

Our first control UI consisted of a Python based GUI with sliders for controlling pitch and yaw, used in conjunction with Skype. We also implemented an Oculus Rift based interface where the remote user would see streamed video on a head mounted display and head orientation controlled the Polly camera view. The Oculus view was not stereo, but seeing the video in an HMD and controlling

[7] https://github.com/umlaeute/v4l2loopback

the camera by head motion felt natural when latency was low, as within a single LAN. However we were concerned that latency over the internet would make it unusable. Also a project priority was easy access to Polly by remote participants, so our subsequent remote interfaces have been exclusively web based.

Our web UI used an HTML5 page with sliders for yaw and pitch, controllable by mouse or arrow keys. The page could be placed next to a Skype window, or in the case of Vidyo or Google hangouts, could be included in the same web page as the video view. After several tests and experience with other users, we found that using Google Hangouts, with our web interface included as a 'Hangouts Extension app' was most expedient. (Figure 3.) We also replaced separate yaw and pitch sliders with a *'panorama view control'* having a draggable view rectangle representing current view within a larger rectangle representing possible views, similar to the interface used in [14]. Potentially in future Polly devices with panoramic cameras, this larger rectangle would show a panoramic image. Our interface also includes a *'bird's eye view'* showing Polly's orientation and current view relative to the guide. Mouse movement in this view can be used to control the Polly yaw axis. Another natural control method that was requested was to have mouse clicks within the Hangouts video window move the camera to recenter on the clicked position. Unfortunately, mouse events are not currently available in the Hangouts API.

Remote participants also requested a way to know where Polly was located and what it was looking at, so for outdoor usage where GPS data is available from the Polly phone, we added a Google Maps interface. The Polly position is shown as a green circle, with a cone indicating the camera heading. Because the streamed video quality is limited by network connectivity, and because taking photos is a natural activity and produces a keepsakes of the experience, we added a 'take-photo' action which captures a full resolution image on the phone, that shows up in the Hangouts app and is stored in a Dropbox folder. Figure 3 also shows that more than one Polly can be used and controlled through our interface.

4.2 Gimbal Usage Modes

The gimbal and SimpleBGC controller provide several operation modes for camera orientation control. One modal distinction is between proportional or differential control, and for the reasons discussed above, we use proportional control. Another modal distinction is between *heading lock* mode and *follow* mode. In heading lock mode, the camera points always in a fixed direction (although this direction may be controlled by the remote operator) *independently* of how the camera is held or the orientation of the frame of the person carrying it. In follow mode, outside of a small "deadband", the camera will gradually orient itself to maintain a fixed angle relative to the carrying frame. Pitch, roll and yaw can all be controlled in either mode, but for pitch and roll, heading lock is nearly always the natural choice and the only mode we support. For yaw however, both modes are useful. Heading lock mode is useful when the remote participant wants to look at something or control their view independently of the motion or orientation of the carrier. Follow mode is useful when the remote viewer wants to

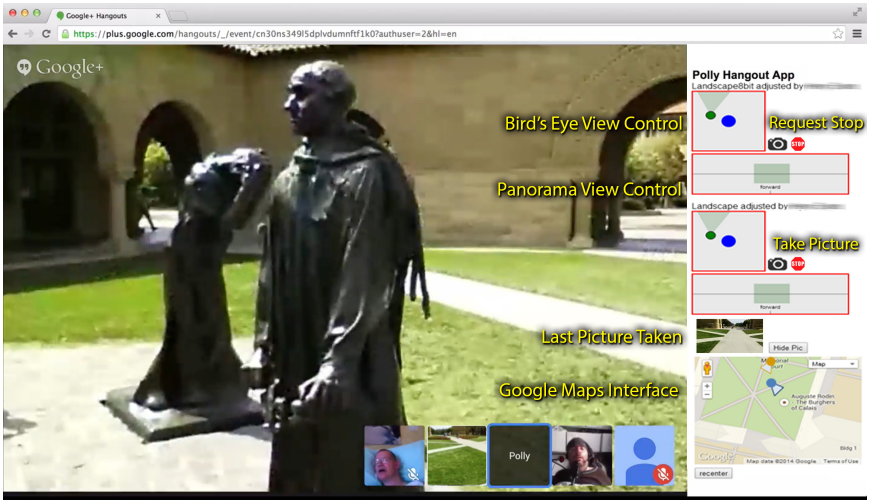


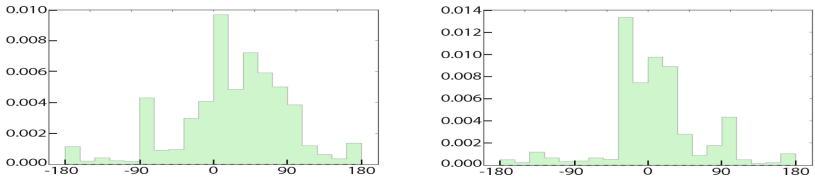
Fig. 3. Polly Hangouts Application

keep looking in the same direction relative to the carrier, e.g. to look forward, matching the forward direction of the carrier, or look sideways to carry on a discussion with him or her.

4.3 "Look at That!" - Interaction Between Local and Remote Participants

A unique aspect of Polly in comparison to many telepresence systems is that the camera view is not only controlled by the remote participants, but also since Polly is mostly operated in follow mode, it is somehow "collaboratively" controlled by the local guide and the remote persons. Thus, the need arises for an easy way to express, even with a bad audio channel, that the remote person would like to keep looking at a particular direction. Another important issue we encountered during all of the tests is to communicate where to look at with Polly. The guide tends to say "Look at that to your left!", but since Polly's camera feed does not provide a good spatial perception of the surroundings and where the guide is looking or perhaps pointing at, it is hard to determine where exactly to look. Usually in our tests this resulted in several instructions back and forth. We recently implemented a new additional interface for the local guide, using Google Glass (See Figure 4). It not only enables the remote operators to request a stop resulting in an alert being shown, but also it displays the bird's eye view, which shows both Polly's and the guide's heading. Future studies are planned to test the effectiveness of the methods described here.

Fig. 4. Google Glass Interface



(a) (b)

Fig. 5. Histograms of Polly yaw relative to frame. (a) Stanford campus tour, (b) Cantor art museum visit.

5 Polly Evaluation and Field Tests

In this section we first describe early experiments with a few common mobile telepresence scenarios using phones or early versions of Polly. These experiments assessed the feasibility of Polly in comparison to just holding a phone. We then describe tests using Polly to provide tours or remote visits for disabled users.

5.1 Initial Phone and Polly Experiments

The intention of the initial experiments was to gain understanding of the benefits of a Polly-style device compared to the base-line of a smartphone running video-conferencing software. These are summarized as the first 10 tests in Table 1. Six of these tests were performed with just a smartphone, and four with a smartphone on a Polly device.

For video communication, Skype was used on seven occasions and Vidyo, Google Hangouts and Ustream were each used once. We found that Ustream provided good video quality, but its high buffering resulted in latency on the order of 30 seconds, making it unusable for remotely controlled scenarios. The phone's internal data connection (all phones were 4G-capable) was used on six occasions and the publicly accessible WiFi at the remote location was used on the other four occasions.

After each of these tests, we asked the remote participants and the Polly guide to jointly fill out an online questionnaire and to provide free-form comments and qualitative statements on a five point Likert scale. Although the sample size is

Table 1. Initial field tests of mobile telepresence scenarios using smartphone or Polly (tests 1-10) and Polly field tests with senior or disabled remote participants (11-16)

Test	Location Type	Scenario	Dur.	Device	Software	Net.
T1	Warehouse club	remote shopping	15	phone	Skype	3G
T2	Costume fair	event visit	60	phone	Skype	3G
T3	Conference	remote attendance	20	phone	Skype	WiFi
T4	Hospital	hospital visit	10	phone	Skype	3G
T5	Aviation museum	museum visit	40	phone	Ustream	WiFi
T6	Lab office building	in-office test	45	Polly v1	Skype	3G
T7	Aviation museum	museum visit	24	phone	Vidyo	WiFi
T8	Computer store	remote shopping	15	Polly v2	Skype	WiFi
T9	Aviation museum	museum visit	45	Polly v2	Skype	3G
T10	Maker faire	event tour	40	Polly v3	Hangouts	3G
T11	Senior center	neighborhood walk	20	Polly v2	Skype	3G
T12	Senior center	remote shopping	25	Polly v2	Skype	3G
T13	Research lab	lab tour	45	Polly v2	Hangouts	3G
T14	Stanford	school tour	60	Polly v3	Hangouts	4G
T15	Stanford campus	campus tour	60	Polly v3	Hangouts	4G
T16	Park	bicycle tour	60	Polly v3	Hangouts	4G
T17	Art museum	museum visit	65	Polly v3	Hangouts	WiFi

too small for meaningful statistical analysis, the results did reinforce some of our observations during the tests. For one, the most negatively rated aspect of the Polly-style interactions was the audio-video connection quality, with a median rating of 2.5. This is also reflected in several user comments, e.g., "*The real number one problem is poor video quality and dropout.*". Another thing we note is that the median physical comfort rating for the guide was higher when using Polly than when holding a phone (4.5 vs. 3.5). The subject in test T1 commented "*As my hands became more engaged with shopping, I wished I had a Polly mount with me!*".

We had some very compelling positive comments, for example: "*I was able to buy exactly what Anya wanted, she was happy to be included in the shopping experience and was happy with the results.*" (T1), and "*It was fun seeing the museum and it was fun being able to look around. I enjoyed being able to switch between looking forward or at objects and being able to look at Steve [the guide].*" (T9).

Reactions from people at the remote locations were generally positive, e.g., "*They seemed amused and interested in seeing Polly. Patrick and Larry had brief polite conversations with Gwen [the guide].*" The guide in test T1 reported, "*People were generally interested and asked about the device*". Another guide commented: "*Got a couple of looks like 'what's that guy doing?'*" (T8).

Table 2. Summary of Stanford Campus and Cantor Art Museum tours

Test	Name	Num Actions	Avg. Dur	Pct. Control	Avg. Yaw	Avg. Pitch	Pct. Collisions
Campus (T15)	Harry	89	10.8s	24.8	99.1	44.2	0.3
	Steve	50	4.3s	5.5	99.2	43.1	2.7
Cantor (T17)	Harry	71	8.4s	13.0	82.1	41.9	0.2
	Jim	6	8.0s	1.0	131.8	36.3	0.0

One behavior frequently observed was that the remote user would move the camera between looking forward in the direction Polly was being carried and looking towards the person carrying Polly. The activity of moving the camera view could sometimes be tedious, but made the experience feel less passive. During discussions involving a few people at the Polly location, the remote viewer would often try to point the camera towards the person talking. This was fairly easy with only a couple of people staying relatively fixed, but could be confusing with more people or when people were moving.

5.2 Tests with Disabled and Elderly Users

Once we were a little further along in the evolution of Polly, we tried to broaden our tests to assess how useful Polly would be as an assistive technology for people with limited mobility.

For our initial investigation, we engaged with the staff and patients in a local senior center. We conducted two field tests as a result of this engagement, T11 and T12. It became clear that while using Polly was interesting to this population, the use of technology that was not familiar presented some additional challenges. Subsequent field tests were with people who had suffered catastrophic events that had caused mobility limiting disability, in some cases as severe as quadriplegia.

The remote participants in field test T13 were a quadriplegic, Henry, who is mute, and communicates primarily by spelling words out to his wife Jane by looking at letters on a letter board. Henry was also able to operate the Polly interface using eye tracking for mouse movements and one finger to click. This test provided us with much useful feedback, such as the suggestion to mount the camera in a landscape orientation, and some comments about usability, that prompted us to develop the Google Hangouts app for controlling Polly. Henry and Jane were also participants in test T15, and were joined by Stuart, who is quadriplegic, but able to speak. Henry and Stuart were able to take turns controlling the position of the camera. Henry remarked “Sharing a Polly turned out to be great fun because of the social aspects - I actually preferred it.”

As the user interactions with Polly had become more complex, we instrumented the interface to collect data about turn-taking. Table 2 summarizes this. Turns are times of continuous adjustment by one user, either through mouse or arrow keys. During test T15, one participant, Henry, took 89 turns, controlling Polly 24.8% of the time, with an average duration per turn of 10.8 seconds. On the average he changed the yaw by 82.1 degrees and the pitch by 44.2. Collisions did not seem to be much problem, as one user trying to get control when another user had control happened less than 3% of the time. The yaw histogram (Fig 5(a)) shows that the most common yaw placement for Polly was forward, but it was also frequently oriented towards the guide.

In a departure from our usual body-mounted configuration, T16 was performed with Polly attached to the handlebar stem of a bicycle, with the guide riding the bicycle. At this point, the Polly Hangouts plugin now contained a map. The remote participants pointed out that the map would be more useful if we could indicate view direction in it, a suggestion we have now incorporated.

T17 took place at the Stanford Cantor Art Museum, with three remote participants, A summary of the adjustment behavior, and histogram of yaw orientation are shown in Table 2 and Figure 5 (b). This test used campus WiFi rather than the phone's data network, and we had more connectivity problems, with five disconnections during the 65 minute session. This test included a museum curator providing expert commentary on the artwork, and a third person not associated with the Polly project as the 'guide' (i.e. carrying Polly.) It highlighted a number of difficulties we still must address. One was interaction between the guide and remote participants. In one case the guide suggested, "Look down at the label under that vase." When the camera did not point down, the guide tried to provide the intended view by leaning forward. However, because Polly compensates for pitch and roll, this had no effect on the camera orientation. A more subtle interaction was when the guide suggested, "Look at the picture to your right. No, further to the right." and then after a few seconds turned his body to point the camera in that direction. Because Polly was in yaw follow mode, this did cause the camera to point in that direction, but can be disorienting to the remote user because the camera is being controlled both locally and remotely. These difficulties suggest that it would be useful to have a simple way for the guide or the remote participant, to request Polly to look in the direction chosen by the guide. Based on the experiences at Cantor museum, the remote participants reported the experience as enjoyable, but it also highlighted the importance of a consistent network connection. It was also suggested that Polly is well suited to outdoor tours, but telepresence robots have some advantages for indoor tours.

6 Discussion and Future Work

Our field tests suggest that Polly type devices can provide an enjoyable mobile telepresence experience, particularly to users restricted by limited mobility, and that the social aspects of Polly use are a positive aspect of that experience.

The biggest current difficulty in using Polly is getting adequate network connectivity to maintain good video and audio quality. However, we expect that over the next few years this will become less of an issue and devices like Polly will find increased use. Furthermore, the capability of taking photos or recording high quality video while streaming at whatever quality is supported by the wireless network allows for the production of high quality video after the fact. Our GoPro based prototype had this capability and we are including it in next Polly version.

Our field trial experience reinforces several hypotheses. These include: (1) Stabilized camera motion is much more pleasing than jerky motion from unstabilized hand held or head mounted views, particularly for bandwidth limited streamed video, (2) The ability to control the camera gives a sense of engagement, even when it is not being exercised, (3) Bad audio can lead to the remote person feeling 'left out' of the experience even when they feel in control of their camera view, and (4) Communication about views and shared camera control between guide and remote participants can be challenging and requires improved user interfaces.

The decision to use a smartphone and commercial streaming apps such as Skype and Hangouts for our current Polly versions allowed us to make it relatively inexpensive and easy to operate, but constrained our use of computer vision. While the streaming apps are running, other apps do not have access to the raw images. Our next Polly version will use Linux with a loopback video device which will allow us to run computer vision algorithms on the video streams, and still use commercial streaming. [8] There are several ways that computer vision could be helpful for Polly. One is the use of QR code or object recognition to provide metadata for museum artifacts. That could also be helpful for localization indoors where GPS is not reliable. Optical flow and tracking could also be helpful for camera control, for example to keep the camera centered on an object of interest selected by the user. Feature matching could also provide better alignment than IMU sensors alone, between the Polly view and a view from Google Glass worn by a guide.

Based on our positive experiences with field trials, we believe devices like Polly will become popular as wireless network coverage improves, processor power increases, and high resolution wide field of view image sensors become available, enabling non-mechanical versions of Polly which use addressable viewports from panoramic cameras.

Acknowledgements. The authors thank Henry and Jane Evans, Adoph Smith, Stuart Turner, Mantvydas Juozapavicius, Steve Cousins, Patience Young and Michael Fischer for participating in trials and providing valuable feedback. We also thank Gwen Gordon, Tony Dunnigan, John Doherty, Susan Roberts-Manganelli and the Fremont High School Robotics team for various contributions and useful discussions.

[8] This may also be possible on our current phone based versions, when a native Android WebRTC client becomes available.

References

1. Ryan, R.M., Deci, E.L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist (1) 68–78
2. Sheldon, K.M., Elliot, A.J., Kim, Y., Kasser, T.: What is satisfying about satisfying events? testing 10 candidate psychological needs. Journal of Personality and Social Psychology **80**(2), 325–339 (2001)
3. Drugge, M., Nilsson, M., Parviainen, R., Parnes, P.: Experiences of using wearable computers for ambient telepresence and remote interaction. In: Proc. 2004 ACM SIGMM Workshop on Effective Telepresence, pp. 2–11. ACM (2004)
4. Mayol, W.W., Tordoff, B.J., Murray, D.W.: Wearable visual robots. Personal Ubiquitous Comput. **6**(1), 37–48 (2002)
5. Mayol-Cuevas, W., Kurata, T.: Tutorial: Computer vision for wearable visual interface. Workingpaperimportmodel: Workingpaperimportmodel University of Bristol Other page information: - Other identifier: 2000803 (2005)
6. Goldberg, K.Y., Song, D., Khor, Y.N., Pescovitz, D., Levandowski, A., Himmelstein, J.C., Shih, J., Ho, A., Paulos, E., Donath, J.S.: Collaborative online teleoperation with spatial dynamic voting and a human "tele-actor". In: ICRA, pp. 1179–1184. IEEE (2002)
7. Tsumaki, Y., Ono, F., Tsukuda, T.: The 20-DOF miniature humanoid MH-2: A wearable communication system. In: ICRA, pp. 3930–3935 (2012)
8. Kashiwabara, T., Osawa, H., Shinozawa, K., Imai, M.: Teroos: a wearable avatar to enhance joint activities. In: Proc. 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 2001–2004. ACM (2012)
9. Adalgeirsson, S.O., Breazeal, C.: Mebot: A robotic platform for socially embodied presence. In: Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction. HRI 2010, pp. 15–22, Piscataway. IEEE Press (2010)
10. Kristoffersson, A., Coradeschi, S., Loutfi, A.: A review of mobile robotic telepresence. Adv. in Hum.-Comp. Int. 2013, 3:3–3:3, January 2013
11. Cousins, S., Evans, H.: Ros expands the world for quadriplegics. IEEE Robotics Automation Magazine **21**(2), 14–17 (2014)
12. Merritt, E.: Center for the future of museums blog, May 2014. http://futureofmuseums.blogspot.com/2014/05/exploring-robots-for-accessibility-in.html
13. Butterill, J.: Virtual Photowalks (2013). http://www.virtualphotowalks.org/
14. Liu, Q., Kimber, D., Foote, J., Wilcox, L., Boreczky, J.: Flyspec: A multi-user video camera system with hybrid human and automatic control. In: Proceedings of the Tenth ACM International Conference on Multimedia. MULTIMEDIA 2002, pp. 484–492, New York. ACM (2002)