Polly: "Being There" through the Parrot and a Guide

Sven Kratz[1], **Don Kimber**[1],
Gwen Gordon[2]
FX Palo Alto Laboratory
3174 Porter Dr.
Palo Alto, CA, 95136, USA
[1]lastname@fxpal.com
[2]gwengordon@comcast.net

Weiqing Su
University of California San Diego
San Diego, California, USA
wsu@ucsd.edu

Don Severns
RCManchild
San Francisco, CA, USA
rcmanchild@gmail.com

Abstract
Telepresence systems usually lack mobility. Polly, a wearable telepresence device, allows users to explore remote locations or experience events remotely by means of a person that serves as a mobile "guide". We built a series of hardware prototypes and our current, most promising embodiment consists of a smartphone mounted on a stabilized gimbal that is wearable. The gimbal enables remote control of the viewing angle as well as providing active image stabilization while the guide is walking. We present qualitative findings from a series of 8 field tests using either Polly or only a mobile phone. We found that guides felt more physical comfort when using Polly vs. a phone and that Polly was accepted by other persons at the remote location. Remote participants appreciated the stabilized video and ability to control camera view. Connection and bandwidth issues appear to be the most challenging issues for Polly-like systems.

Author Keywords
telepresence, remote guiding, wearable, gimbal, user feedback

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Figure 1: The current Polly prototype consists of a smart phone mounted on a three-axis brushless gimbal that can be carried, placed on perches or flat surfaces, or worn by the guide at shoulder level.

Introduction

Two fundamental human traits are the curiosity to explore our environment and the need to be close to loved ones. 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 even rock 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 view of the person carrying it[1]. Polly consists of a three-axis stabilization gimbal driven by silent brushless motors holding a mobile phone that provides the audio and video feed and a connection to the internet. Polly can be carried by hand, placed on "perches", rest on surfaces, or worn by way of a backpack frame with an attachment holding Polly near the shoulder of its wearer.

We implemented an application that lets the remote user control the gimbal orientation (pitch and tilt) from their location. We used Skype[2], Vidyo[3] and Google Hangouts[4] for audio and video transmission, but any suitable software with a mobile client should work with Polly.

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.[5] 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 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 brushless gimbal uses IMUs for active rotation compensation and leveling to produce a very smooth video feed, much like a steadicam rig, but much smaller in size and weight.

In the rest of this case study, we describe the design process of Polly from its initial prototype to its current shape and contribute a set of initial design observations based on several field deployments of Polly we have conducted in the San Francisco Bay Area, USA and Cleveland, OH, USA.

[1] Also motivated by the remarkably stabilized head pose of many birds. http://www.youtube.com/watch?v=UytSN1Hw8J8

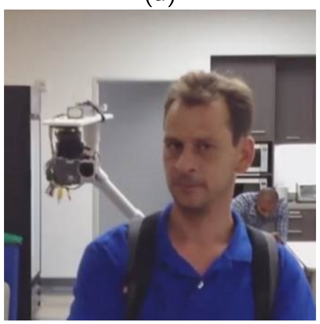
[2] http://www.skype.com

[3] http://www.vidyo.com

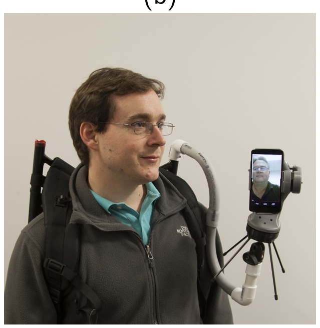
[4] http://www.google.com/hangouts/

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

(a)



(b)



(c)

Figure 2: Polly Prototypes—
(a) first frame mounted version,
(b) first stabilized version,
(c) current prototype.

Related Work

Drugge et al. proposed a wearable telepresence system consisting of an HMD and a head-mounted camera [1]. In contrast to Polly, this does not allow the remote participant the same degree of control, as view direction cannot be changed, as is the case using Polly's gimbal. Similarly, systems such as Google Glass streaming to a hangout, or the Tele-actor system [2] do not give the remote user any direct control over view. The MH2 [4] is a shoulder-worn humanoid telepresence robot with a focus on conveying gestures and poses made by the remote participant. Polly, on the other hand is not based on embodiment by physical shape. Rather, a camera image of the remote participant, displayed via a smartphone, is used to embody him or her. *TEROOS* [3] is a shoulder-mounted wearable telepresence system, that is perhaps the wearable telepresence system with the closest resemblance to Polly, as it is also uses a gimbal that is 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, makes for very low quality videos while the local operator is walking. Secondly, TEROOS follows the avatar concept, similar to MH2, and uses an abstracted form of embodiment in the shape of a decorated camera. Again, we believe that actually showing the remote operator directly may be advantageous, for example in cases where the mobile user knows the person he is interacting with personally.

Development of the Polly Prototypes

The baseline case for Polly 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 hangout. 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.

First frame mounted version

The first 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, some digitization hardware, and a battery. We created two versions, one using a Logitech 920C webcam, connected to a Microsoft surface running Skype. Another version used a GoPro which could record HD video onto a SD card, while simultaneously outputting analog video, which was digitized using a USB digitizer. We tried several vendors for USB digitization, including Diamond and Hauppauge. None of them worked well directly as the video input for Skype, but we found that using third party virtual camera programs such as XSplit[6] or ManyCam[7], the video from the GoPro could be input to Skype to be streamed.

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 steady cam type stabilization, and for which the camera can be pointed remotely. The gimbal used was a three axis brushless motor gimbal designed for a GoPro camera on a

[6] http://www.xsplit.com/
[7] http://www.manycam.com/

small UAV, and using the Alex Mos SimpleBGC[8] board and software. These kind of gimbals have become very popular over the past year and a half for UAV use, and are starting to be used for hand held camera stabilization. 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.

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 with the Diamond device. (GoPro also outputs HDMI, but we could not find a good portable solution for making HDMI video available as input to Skype.) One difficulty is that wires carrying the video signal from the cameras must cross three stabilization motor axes, so its 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 these problem with slip rings, but found problems with high frequency noise with the slip rings.

Current versions

The biggest difficulty we had with the first stabilized version was the complexity of the system requiring the camera, gimbal, multiple USB devices, and MS Surface tablet, in addition to intermediate programs such as XSplit that were necessary in addition to the basic video conferencing program. We decided the best approach for a much simpler more usable device would be an entirely self-contained unit consisting only of the gimbal and a smart phone (Figures 1 and 2 (c).) In this design, the camera, videoconferencing software, and all necessary control software is provided on the phone. The gimbal case contains a battery, and a Bluetooth 4.0 BLE to PWM device which the user never needs to be aware of. That allows a PollyControl App running on the phone to receive messages sent from the remote user, and send the necessary PWM signals to the gimbal via the BLE device. Another advantage of the current 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. We are also in the process of designing additional mounting options, including a "perch" which can provide power, allowing permanent operation.

User Interface and Usage Modes

We implemented several versions of user interface for the remote viewer to control Polly. The simplest consists of a wxPython based GUI with sliders for controlling pitch and yaw, which can be used in conjunction with Skype. We also implemented an Oculus Rift based interface where the viewer would see the streamed video on a head mounted display, and head orientation would control the direction of the Polly camera. Most of our usage however has been with a web based interface that allows sliders or arrow keys for controlling the view. We experimented with several usage modes for the prototypes involving stabilization gimbals. One modal distinction is between *angle mode*, in which the yaw or pitch angles are proportional to slider position, and a *speed mode* in which the sliders specify angular velocity. For low latency, the speed mode is easy to use, but when the latency is more than about a half second, it becomes difficult to control. Direct control angle mode works much better as the latency increases.

[8] `http://www.basecamelectronics.com/simplebgc/`

Test ID	Location Type	Region	Scenario	Duration (min)	Device	Software	Connectivity
T1	lab office nuilding	R1	in-office test	45	Polly	Skype	phone data connection
T2	aviation museum	R1	museum visit	60	Polly	Vidyo	location's WiFi
T3	computer store	R1	remote shopping	15	Polly	Skype	location's wiFi
T4	aviation museum	R1	museum visit	45	Polly	Skype	phone data connection
T5	warehouse club	R1	remote shopping	15	phone	Skype	phone data connection
T6	historic costume fair	R1	event visit	60	phone	Skype	phone data connection
T7	conference	R1	conference attendance	20	phone	Skype	location's WiFi
T8	hospital	R2	hospital visit	10	phone	Skype	phone data connection

Table 1: Summary of field the field tests of Polly-style remote interactions conducted so far. Region R1 is the San Francisco Bay Area in California, USA, and Region R2 is the Cleveland Metro Area in Ohio, USA.

Another modal distinction is between *heading lock* mode and *follow* mode. In heading lock mode, the camera points 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 so that its yaw is relative to the carrying frame. The heading lock mode is useful when the remote participant wants to look at something independently of the motion of the carrier. Follow mode is useful when the remote viewer wants to look forward (matching the forward direction of the carrier), or look sideways to carry on a discussion with the carrier.

Preliminary Evaluation of Polly

To gain further insight into Polly-style usage scenarios we conducted a total of 8 tests at different locations. Four tests were done with Polly, and four with just a smart phone running a video conferencing software. One reason for this is that we wanted to see how Polly would compare with a less sophisticated solution. Also, some situations without using Polly simply arose spontaneously. As shown in Table 1, studies were conducted at a total of 7 distinct locations, in two different regions of the USA. For video calls Skype was used on 7 and Vidyo on one occasion. The phone's internal data connection (all phones were LTE-capable) was used on 5 occasions and the remote location's publicly accessible WiFi was used on 3 occasions.

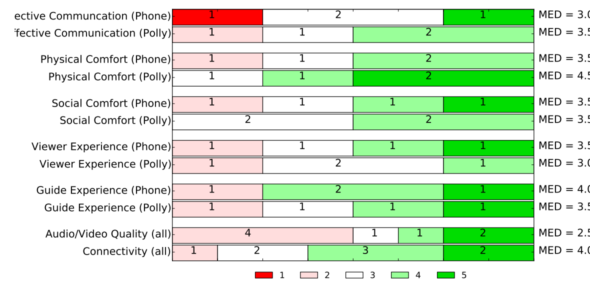


Figure 3: Qualitative feedback obtained from post-test questionnaire. Items were rated on a 5 point Likert scale.

Qualitative Feedback
After each (formal or spontaneous) test, we asked the remote participants and the Polly guide to jointly fill out an online questionnaire with the data represented in Table 1 and, in addition, free-form comments and qualitative statements on a five point Likert scale. Figure 3 shows an overview of the this feedback. Although the sample size so far is too small to do meaningful statistical analysis, Figure 3 does highlight some issues we perceived during the tests. For one, the most negatively rated aspect of the Polly-style interactions was the audio quality, with a median rating of 2.5. This is also reflected in several user comments, e.g., Don commented: "*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 (4.5 vs. 3.5). Dave made the following comment on this aspect: "*As*

my hands became more engaged with shopping, I wished I had a Polly mount with me!".

Other Observations
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. Having presets for forward and for eye-to-eye engagement between local and remote user might be useful. During discussions involving a few people at the Polly location, the remote viewer would often point the camera towards the person talking. This was was fairly easy with only a couple of people staying relatively fixed, but could be confusing with more people or when people moved more.

Discussion and Future Work

The single biggest difficulty we had in using and testing the system was getting a consistently high enough bandwidth connection 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 recording high quality video on Polly while streaming whatever lower quality is supported by the wireless network allows for the production of a high quality video after the fact. Our earlier GoPro based prototypes had this capability, and we expect to add it to our future versions.

Our personal experience and the limited initial tests suggest several hypotheses. These include: (1) Stabilized camera motion is much more pleasing than jerky motion from unstabilized hand held or head mounted views, (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.

Based on these results, we plan to next identify specific interesting locations that provide consistent wireless network, and do more careful studies based on larger numbers of users, particularly drawn from special populations with mobility issues such as seniors, or the disabled.

Acknowledgments

Gimbal technology and design work for this project were provided by RCManChild. (`http://rcmanchild.com/`)

References

[1] Drugge, M., Nilsson, M., Parviainen, R., and Parnes, P. Experiences of using wearable computers for ambient telepresence and remote interaction. In *Proc. 2004 ACM SIGMM workshop on Effective telepresence*, ACM (2004), 2–11.

[2] Goldberg, K. Y., Song, D., Khor, Y.-N., Pescovitz, D., Levandowski, A., Himmelstein, J. C., Shih, J., Ho, A., Paulos, E., and Donath, J. S. Collaborative online teleoperation with spatial dynamic voting and a human "tele-actor". In *ICRA*, IEEE (2002), 1179–1184.

[3] Kashiwabara, T., Osawa, H., Shinozawa, K., and Imai, M. Teroos: a wearable avatar to enhance joint activities. In *Proc. 2012 ACM annual conference on Human Factors in Computing Systems*, ACM (2012), 2001–2004.

[4] Tsumaki, Y., Ono, F., and Tsukuda, T. The 20-dof miniature humanoid mh-2: A wearable communication system. In *ICRA*, IEEE (2012), 3930–3935.