

HomeProxy: Exploring a Physical Proxy for Video Communication in the Home

John C. Tang¹, Robert Xiao^{1,2}, Aaron Hoff¹, Gina Venolia¹, Patrick Therien¹, Asta Roseway¹

¹Microsoft Research

One Microsoft Way

Redmond, WA 98008

{johtang|aaronho|ginav|patherie|astar}@microsoft.com

²Carnegie Mellon University

5000 Forbes Avenue

Pittsburgh, PA 15213

brx@cs.cmu.edu

ABSTRACT

HomeProxy is a research prototype that explores supporting video communication in the home among distributed family members through a physical proxy. It leverages a physical artifact dedicated to representing remote family members to make it easier to share activities with them. HomeProxy combines a form factor designed for the home environment with a “no-touch” user experience and an interface that responsively transitions between recorded and live video messages. We designed and implemented a prototype and conducted a pilot study with eight pairs of users. Our study demonstrated the challenges of a no-touch interface and the promise of offering quick video messaging in the home.

Author Keywords

Video chat; home; physical proxies; asynchronous video.

ACM Classification Keywords

H.5.2. Information interfaces and presentation : Prototyping

General Terms

Human Factors; Design.

PHYSICAL PROXIES FOR VIDEO CHAT AT HOME

Several consumer tools for video communication, such as Skype, Apple’s FaceTime, and Google+ Hangouts, have made video chat in the home commonplace. People have been using these video tools to share diverse experiences among family members and friends, such as cooking, watching TV, and celebrating holidays [1]. Research prototypes have also explored live, always-on video connections to increase connectedness among distributed family members [6]. The Skype Cabinet [7] and ShareTable [10] prototypes have even explored an implicit interface for starting live video connections by opening a cabinet door.

Meanwhile, research has demonstrated the advantages of physical proxies for representing remote workers participating via video conferencing [9]. Giving remote video colleagues a physical embodiment enables smooth

turn-taking, enhances their participation in meetings, and, over time, integrates them socially into the team. Using proxies for temporally remote work colleagues through recorded video [6] has also been explored. HomeProxy leverages these advantages of a physical proxy in sharing activities via video in the home context. We expect that a dedicated, physical representation of a remote family member will allow more naturally sharing activities with them wherever they occur throughout the home.

Designing for the home context involves tailoring the aesthetics and a number of important user experience factors. The scope of potential users, from kids to the elderly seniors, defines a wide range of user capabilities for operating the interface. Remote family members may span across different time zones [3], and people’s availability in the home may be less predictable than in the workplace. These temporal factors suggest that asynchronous video messaging may be useful and popular for staying socially connected in the home [5]. While the Pêle-Mêle project [4] explored novel interfaces to live and recorded videos in the home, they used abstract layouts and representations on a wall-mounted display, whereas HomeProxy aims to be a portable, physical proxy for a geographically or temporally remote person.

The HomeProxy prototype explores connecting distributed family members through a physical proxy that supports both synchronous and asynchronous video communication. We describe the design goals of HomeProxy, a prototype we implemented, initial experiences with it in a pilot user study, and plans for future work.

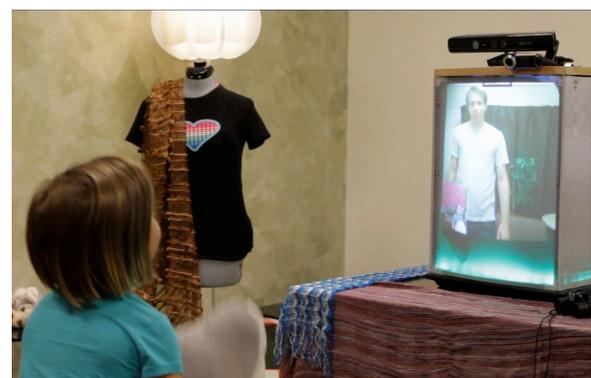


Figure 1. HomeProxy transmitting a live video call.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI '13, April 27 – May 2, 2013, Paris, France.

Copyright 2013 ACM 978-1-XXXX-XXXX-X/XX/XX...\$10.00.

HOMEPROXY DESIGN GOALS

HomeProxy was driven by three design goals: 1). designing a video communication physical proxy for the home environment, 2). experimenting with a “no-touch” user experience, and 3). exploring the transition between recorded and live video communication.

Design for the Home

HomeProxy was designed to be an appliance that looks a bit like a lamp to fit into the home aesthetic (Figure 1). A slightly bowed rear-projected screen, fabric sides, and wood top were used to evoke more of a consumer feel. Its portable and self-contained package affords moving among the various locations (e.g., dining table, living room couch, child’s bedroom) where family activities occur in the house.

“No-touch” User Experience

Given the wide range of users in the home environment, we experimented with a “no-touch” user experience that did not involve any buttons or user interface controls. HomeProxy “woke up” automatically when a user approached and “went to sleep” when a user left. The device itself was always on to sense for the proximity of users in the environment, but from the user’s perspective, it activated in response to their presence.

HomeProxy was controlled by a small set of simple, natural gestures. This goal was achieved by focusing on just three functions: recording a video message, playing back a video message, and breaking-in to a live, synchronous video chat while someone was recording a video message. This user experience did not require manipulating any user interface device and also avoided unsightly smudges that have become an eyesore in many touch-based devices, especially in home environments.

Recorded and Live Video Transitions

While synchronous video chat has become quite popular and research prototypes have tried asynchronous video messaging [5], we wanted to explore transitioning between recorded and live video communication. While we expected that the home setting will benefit from the flexibility of recorded messaging, we also wanted to enable live chatting if both sides happen to be available. This combination of live and recorded video goes beyond prior work that focused just on either live [7, 10] or recorded [5, 8] video.

Simplifying Assumptions

Our iterative design process to develop and implement the first HomeProxy prototype started with some simplifying assumptions. Each HomeProxy connected with only one other HomeProxy to form a dedicated link between two homes. We did not allow specifying an individual within a home, but assumed that messages can be shared within a household without privacy concerns. HomeProxy evokes a metaphor of the traditional home phone line with an answering machine in that calls and messages to that line can be received by anyone in the home.

INTERACTING WITH HOMEPROXY

When a user approaches HomeProxy, it automatically wakes up and presents a home screen (Figure 2). A filmstrip at the top shows any video messages available to view. Each message animates through a sample of still frames from the message, providing a preview to help the user select which one to play. Yellow highlights indicate new messages that have not yet been viewed. The filmstrip can be positioned underneath the selector, outlined in black, by a swipe gesture to the left (right arm swipes left) or right (left arm swipes right). Once the desired video message is positioned within the selector, it begins to play.



Figure 2. Home screen for starting a session.

The glowing LED lights around the base of HomeProxy give an indicator of presence at the remote site [1] so the user can expect whether to have a recorded or live video interaction. When the LEDs glow red (as in Figure 2), there is no one near the remote HomeProxy. If someone is near but not looking at the remote HomeProxy, the LEDs glow yellow, and if someone is looking directly at it (perhaps viewing a video message), the LEDs glow green. Changes in state were also accompanied with ambient sounds.

To initiate communication with the remote site, the user waves with two hands, similar to getting someone’s attention in face-to-face interaction. The wave initiates recording a video message onto the remote HomeProxy. While recording the video message, the remote HomeProxy plays the audio being recorded, animates through selected frames being recorded, and flashes the LEDs blinking green. The recording can be ended by repeating the wave gesture, or by simply leaving the proximity of HomeProxy.

However, if someone is at the remote HomeProxy while a message is being recorded for it (indicated by the incoming audio, animating frames, and blinking green LEDs), that person can break-in to a live call through the two-handed wave gesture at their HomeProxy. This causes a ringing sound and a circulating green LED pattern at the originating

HomeProxy before a live video call is connected. The live call is ended by either side waving or leaving proximity with HomeProxy. User interactions with HomeProxy are illustrated in the accompanying video figure.

BUILDING A PROTOTYPE

Realizing the design goals for HomeProxy into a prototype involved combining several different software components to create an integrated, no-touch user experience into a hardware prototype. The system diagram for two connected HomeProxies is shown in Figure 3.

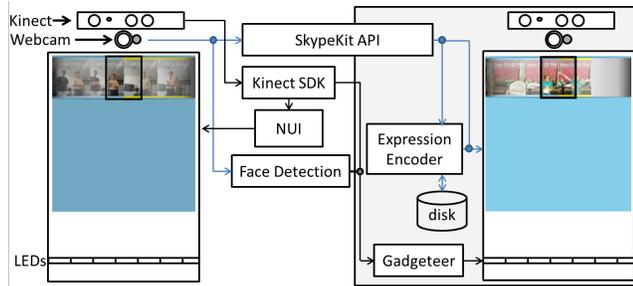


Figure 3. System diagram for connecting two HomeProxies.

The SkypeKit API was used to transmit video and audio captured by a webcam (Logitech Pro 9000) between the HomeProxies. A continuous Skype call is initiated when first starting HomeProxy, and that call is suspended and resumed depending on user interaction with HomeProxy. Recording a video message actually streams the video across the Skype call to the remote HomeProxy where it is saved on a local disk using Expression Encoder. This arrangement substantially reduces the startup latency of recording or viewing a message or breaking-in to a live call, making the system highly responsive. A virtual Skype video renderer was written in C++ to enable the video stream opened by SkypeKit to be used concurrently for the face detection and video windows in the interface.

Kinect was used via the Kinect SDK for Windows to detect user proximity. Kinect detected if there was a user in the vicinity of HomeProxy, to distinguish between the red and yellow LED states. A face detection algorithm ran on the video stream from the webcam to detect if a user was actively looking at HomeProxy, to distinguish between the yellow and green LED states. The software used a programmable Gadgeteer microcontroller (FEZ Hydra) to interface with a strip of addressable LEDs (Adafruit).

Kinect was also used to implement the no-touch user interface. Left and right swipe gesture commands that were pre-packaged with the Kinect SDK were used to position the video filmstrip under the selector. A gesture recognizer written in C# was used to detect the two-handed wave gesture for starting and ending video communication.

All the components were fit into a custom enclosure that included a picoprojector (LG HX350) aimed at a slightly bowed optical diffuser surface. An auxiliary lens mounted on the projector and two first surface mirrors for folding the

optical path generated a large projection image while still fitting in the compact enclosure. A USB speakerphone (Polycom CX100) was used for the audio. Diaphanous fabric side panels and a wood top with handle gave it a home furnishings look. Our initial prototype was not self-contained, and was attached to a desktop computer (Core 2 Duo, Windows 7) via a number of USB and VGA cables and connected to the network. Based on our experiences with the prototype, we expect that the computational needs could be handled by a compact slate computer mounted inside the enclosure and networked via Wi-Fi.

PILOT TESTING THE PROTOTYPE

Once we had a working prototype, we did some pilot testing to give us design feedback for the next iteration of HomeProxy. While we only constructed one custom enclosure, we created a functionally equivalent prototype from standard components (computer connected with LCD display, Kinect, speakers, and LED strip) for pilot testing.

We recruited eight pairs of participants (9 female, 7 male) from our work organization and their families. Most of the pairs were friends with each other, and one pair included a mother and her 9 yr. old daughter. After getting a demonstration of HomeProxy, they were then separated into two rooms, each with its own HomeProxy. One person was asked to think of an information sharing or seeking task to share via HomeProxy (e.g., a recent trip or event they experienced or were planning, a favorite activity or local place to visit). They exchanged a few recorded messages on the topic, culminating in the person breaking in to have a live video chat. After that, the participants switched rooms and roles, so that each person had the experience of using the custom enclosure prototype and breaking in to a live chat. A survey documented their perceptions on how easy and quick it was to record and view video messages, what feature they most liked, and what they thought most needed improvement. Sessions typically lasted on the order of a half hour, and each participant received a nominal gratuity (lunch voucher in the company cafeteria) for their time.

Usability Reactions to the No-touch Interface

The pilot studies quickly identified the challenges in doing the no-touch user experience in a robust way. While we had intended HomeProxy to naturally react to the gestures used in communication (e.g., waving to get someone's attention and to end a communication), our prototype showed the practical problems involved in developing such an interface. While the pre-packaged swipe gestures were fairly reliable, as they were trained using a substantial amount of machine learning data, the two-handed wave recognizer that we developed was more problematic. It had a difficult time recognizing all the different ways in which people made a two handed wave gesture and was subject to false positives when people waved during the course of their recorded or live video interactions, causing HomeProxy to terminate the video connection. In the

surveys, 7 out of 16 mentioned some aspect of the gesture control as the feature most needing improvement.

Given the uncertainty around the gesture recognition, users ironically reacted by being more constrained, feeling like they had to stand straight and directly in front of the HomeProxy a certain distance away from it and trying to conform their gestures to what they postulated it could recognize. Their reactions pointed out the need to provide more feedback on the gesture recognition process to help the users successfully enact the gesture commands.

Furthermore, users reduced the expressiveness of their gestures during the video interactions for fear of inadvertently triggering the gesture command to end the video. While the promise of natural user interfaces that Kinect enables is intriguing, our initial prototype was not robust enough to deliver on that promise, and ended up restraining users in their interactions with HomeProxy.

Perceptual Reactions to HomeProxy

Most users liked the video affordance of HomeProxy, with 9 out of 16 rating some aspect of video as their favorite feature. Compared to voicemail, texting, or email, five users preferred HomeProxy videos (e.g., more convenient, easy, natural) and one appreciated the additional context video added over a disembodied voice. People particularly liked how it enabled quick access to video messaging, as four people found it faster than text (either messaging or email). This responsiveness was enabled by dedicating HomeProxy to connect with one paired system, automatically activating when a user approached, and the low latency video chat architecture we employed so a video recording or call started almost immediately. This quick access to video messaging may enable new uses of video in the home.

Several also mentioned that they liked the home look and feel of HomeProxy (the fabric enclosure, the glowing lights, the sounds—even if they did not understand what they meant). While all the participants were able to complete the tasks of recording and viewing messages and breaking in for a live chat in the context of a pilot study, we need to study how useful these features are in actual home use.

FUTURE WORK

After a design iteration to address the main usability features identified in the pilot, we want to deploy HomeProxy into people's homes to see how they would use it in communicating with their family and friends. Since users seemed to be most interested in the video aspects of HomeProxy, we are focusing our design goals for this next iteration to gain use experience with the video features in everyday living. Toward that end, we are deferring on the no-touch interface and re-designing HomeProxy to work on a slate computer form factor. A slate offers a much more portable form factor while still having enough computing power to handle the video communications. We will rely on users' familiarity with touch interfaces that have become

popular in consumer devices. This approach will give us more flexibility in the wider range of user interactions we can support. While we are getting this usage experience, we will be iterating on the design of a more compact hardware enclosure and exploring ways of re-designing the no-touch interface to be more robust through machine learning.

We especially want to learn from a home deployment about the usefulness of transitioning between live and recorded video messages and the benefits of quick access to video messaging. We want to use HomeProxy to help us learn how to blur the line between recorded and live video communication and how that will enable new ways to share social experiences with remote family and friends.

ACKNOWLEDGMENTS

We thank all the anonymous participants who volunteered for our pilot studies. We also acknowledge our colleagues Xiang Cao and Kris Crews who helped with this project.

REFERENCES

1. Bentley, F., K. Tollmar, D. Demirdjian, K. Koile, T. Darrell, Perceptive presence, *IEEE Computer Graphics and Applications*, 23(5), 2003, 26-36.
2. Brubaker, J. R., G. Venolia, J. C. Tang, Focusing on shared experiences: Moving beyond the camera in video communication. *DIS 2012*, 96-105.
3. Cao, X., A. Sellen, A. J. Brush, D. S. Kirk, D. Edge, X. Ding, Understanding family communication across time zones. *CSCW 2010*, 155-158.
4. Gueddana, S., N. Roussel, Pêlè-Mêlè, a video communication system supporting a variable degree of engagement. *CSCW 2006*, 423-426.
5. Inkpen, K., H. Du, A. Roseway, A. Hoff, P. Johns, Video kids: Augmenting close friendships with asynchronous video conversations in VideoPal. *CHI 2012*, 2387-2396.
6. Judge, T. K., C. Neustaedter, S. Harrison, A. C. Blose, Family portals: Connecting families through a multifamily media space, *CHI 2011*, 1205-1214.
7. Rose, D., Ambient Furniture Overview, <http://vimeo.com/40842568>, verified Dec. 19, 2012.
8. Tang, J. C., J. Marlow, A. Hoff, A. Roseway, K. Inkpen, C. Zhao, X. Cao, Time travel proxy: Using lightweight video recordings to create asynchronous, interactive meetings. *CHI 2012*, 3111-3120.
9. Venolia, G., J. C. Tang, R. Cervantes, S. Bly, G. G. Robertson, B. Lee, K. Inkpen, Embodied social proxy: Mediating interpersonal connection in hub-and-satellite teams. *CHI 2010*, 1049-1058.
10. Yarosh, S., A. Tang, S. Mokashi, and G. D. Abowd, "Almost Touching:" Parent-Child Remote Communication Using the ShareTable System, *CSCW 2013*, in press.

