Sharing Experiences over Video:
Watching Video Programs together at a Distance

Anna Macaranas[1,2], Gina Venolia[1], Kori Inkpen[1], and John C. Tang[1]

[1] Microsoft Research, One Microsoft Way, Redmond, WA 98008 USA
[2] School of Interactive Arts and Technology, Simon Fraser University,
Surrey, BC V3T 0A3 Canada
`amacaran@sfu.ca`,
`{ginav,kori,johntang}@microsoft.com`

Abstract. While video communication is becoming quite popular among remote friends and family, recent usage practices have been extending beyond just talking heads to remotely sharing an experience by doing an activity together. However, current video chat tools are aimed at sharing talking heads and need to be reconsidered to support remotely sharing activities. We explore a specific remote shared activity – watching video programs – through a three-phase study. We surveyed people's interest in watching video together, studied how people currently watch together in their homes, and compared different conditions for watching together in the lab. Our work explored people's current and desired practices, interactions, and technical implementations. We present our findings in themes that provide insights for designing systems that better support using video-mediated communication to share watching videos together over distance. We found that remotely watching video programs together while connected by video-mediated communication is engaging, fun, and fosters social bonds between the participants, and that these results are stronger with increased fidelity of the communication media.

Keywords: Shared experiences, proxy, telepresence, teleconferencing, video conferencing, video-mediated communication, home.

1   Introduction

Recent trends in technology have made video-mediated communication (VMC) widely available for family and friends to maintain long distance relationships. In an exploratory study on how people use VMC to stay connected, Brubaker *et al*. [6] found that distributed family and friends used video in tandem with activities such as cooking, watching TV, and even giving birth. In a similar study, O'Hara *et al*. [25] found that mobile VMC was being used for casual conversation, show-and-tell, and everyday routines (i.e., entertaining one's child during bath time). We are intrigued in how people are appropriating current VMC technologies to go beyond "talking heads" conversations (which can be awkward to sustain for prolonged periods of time), to shared experiences that range from everyday routine events to major life events

[6,18,20,25]. Prior work included stories of awkwardly chasing a toddler with a laptop to share with grandparents or snaking a webcam into an engine compartment to diagnose a problem with a father. This need for appropriation identifies ways that current technologies do not directly support the use of video to enable sharing an activity (other than talking) together over distance. We wanted to explore how to design systems that mitigate the issues inherent in VMC (i.e. background noise, privacy, framing – see [18,25]) and foster doing an activity together with someone in a different location, i.e. to remotely share an experience.

We focus on watching video together across distance as a specific kind of remote shared experience. Watching movies, TV shows, or other online video content is a popular social activity. Yet, with the dispersion of friends and family across different geographic locations, users might be remote from the people with whom they would like to watch together. We wanted to explore how adding watching videos together to VMC can use video to naturally sustain sharing activities together. This approach contrasts with adding communication to enable social TV watching over distance, which has been extensively explored in prior work [7,9,24,26]. We use watching video together not only as a remote shared experience to design for, but as a specific activity to learn more about remote shared experiences in general. One particular question we wanted to investigate was whether the combination of audiovisual entertainment with audiovisual communication would result in a divided-attention task in which neither activity was pleasant or effective.

In this paper, we describe a three-phase study that explores how people use VMC (e.g., Skype, Google+ Hangouts, FaceTime) in tandem with remotely watching videos together to create a shared experience. We present the concepts behind shared experiences and the relevant literature that has explored this area within Human Computer Interaction. We describe our three-phase methodology that integrates a survey, exploratory field study, and comparative lab experiment. We analyze the results of our study to provide guidelines on how to design systems to better support video communication during remote shared activities.

2   Related Work

2.1   Beyond Talking Heads to Shared Spaces

Recent work shows that VMC has multiple benefits when the field of view shifts from "talking heads" and towards visually sharing an environment [19,22,27]. Research has shown that shared visual environments can improve communication [22,27], task performance [22] and help establish common ground [8]. Common ground is the mutual awareness between a group of people in understanding the state of the task, as well as another member's state, intention, and comprehension in relation to the task [8]. By achieving common ground, group members can be more focused on completing the task as opposed to coordinating each other's roles in regards to it.

2.2 Shared Experiences

While much work has explored sharing visual environments for collaborative work, a recent wave of work has explored ways in which shared spaces and VMC can create shared experiences in home and consumer settings. Previous work has shown a desire for geographically separated family and friends to stay in touch and be part of the day-to-day activities [6,18,20,25]. Unlike the work focused on understanding how VMC improves collaborative tasks, this wave of work is focused on how it can help maintain and improve the social relationships between people. We expand this work by focusing on watching videos together as a shared activity domain.

2.3 Leisure Activities and Social TV

In our research, we focus on how pairs of remote individuals share leisure activities. Previous work has shown that regular participation in leisure activities has positive effects on physical and mental health [13] and dyadic relationships [2]. In this study, we focus on people who are remotely watching video programs together -- watching video programs (e.g., broadcast TV, recorded movies) together but from different locations. We chose this activity because it is a good example of a leisure activity that can easily be done remotely. When synchronized, the video program can act as a shared space and create a common ground for conversation.

In recent years, many researchers have built systems that embed a communication channel into the TV console (i.e. SocialTV or iTV) [7,9,24,26]. While much of the research has focused on building systems that share audio or text messaging while remotely watching TV together, some research has focused on the social interaction in these settings [11,12,15,23]. Dezfuli *et al.* found that people most often gathered together locally to watch video programs with family but had a high preference for watching with close friends [11]. Ducheneaut *et al.* found that the social discussion among the people was largely shaped and carefully crafted to fit in with the flow of the TV program, both when watching collocated and when watching remotely with an audio connection [12]. Harboe *et al.* ran field studies to see if people enjoy using audio-based SocialTV, how they behaved and what problems they faced doing this activity [15]. They found that people valued the experience, even if reluctant to try it at first. Metcalf *et al.* found that people felt constrained by pre-set text messages and wanted richer forms of communication while watching video programs [24].

We expand on Dezfuli *et al.* and Ducheneaut *et al.'s* work on watching video programs locally by testing their findings on relationship strength and conversational patterns on people watching video programs remotely. We use the insights found from Metcalf *et al.* and expand on Harboe *et al.*'s work by running a field study that uses *video* chat as the means of communication. While Harboe *et al.* speculate that having video chat compete with the TV program for visual attention will be problematic in SocialTV, no research has been done to validate this speculation. Unlike Harboe *et al.*, we predict that people will enjoy video chat while watching video programs.

Our research focuses on enhancing the experience of video communication by augmenting it with the leisure activity of watching videos together. This approach complements most of the work on SocialTV that enhances the leisure activity of TV watching by augmenting it with communication. We build on Flora and Segrin's exploration on the effect of different leisure activities on relationships [14], by focusing specifically on watching video programs together and better understanding how to support it. While Aaltomen *et al.* [1] used groups of strangers to study how VMC configuration affected the experience of a remote shared activity, we use groups of people with an established relationship and who enjoy spending time together.

3   Methodology

In this section we describe the three studies comprising this paper. We combine our findings of the three phases and present the results at the end based on key themes we identified that were consistent throughout the phases.

3.1   Initial Survey

To contextualize our studies, we wanted to understand the relevance of watching video together in people's everyday lives. In particular, we had two questions in mind:

- Are people watching video programs together remotely or interested in doing so?
- If people are participating in this activity, how are they doing it?

We began exploring the answers to these questions with a survey. We ran this survey across two groups: full time employees and interns in our global software engineering company. A random selection of 1000 full time employees was invited to do this study via e-mail. All interns working during the summer received an invitation through a distribution e-mail list. 106 participants (59 full time, 47 intern) completed the survey. The majority of participants were male (78% in the full time group, 64% in the intern group). The majority of full time employees were in their 30's (49%) and 40's (32%). Almost all intern students were in their 20's (94%).

The survey consisted of 37 questions separated into five sections: basic demographics, current practices, desired practices, desired interaction, and miscellaneous. In basic demographics we collected age, gender, video program watching habits and video chat habits. Most importantly, we asked if they had ever communicated with someone at a different location while watching the same video program. This determined which sections they needed to complete. The number of responses to any question varied due to this branching and because all questions were optional.

Participants who had tried to watch video programs remotely completed the current practices section. It explored what they watched, with whom, what devices they used to watch video programs, and what devices and applications they used to communicate. Participants who had attempted or wanted to watch video programs remotely completed the desired practices and desired interaction sections. The desired practices section was almost identical to current practices, allowing us to compare what

participants wanted versus what they did. In desired interaction, participants ranked which forms of interaction (e.g., touch, sight, talking) were most desirable as well as the best configuration to see and hear someone from a different location. All participants completed the miscellaneous section. Here, participants were able to share their thoughts and experiences regarding watching video programs remotely. The survey took approximately ten to fifteen minutes to complete.

3.2   Field Study

While the survey was effective in gauging how many people were interested in or regularly doing this activity and how they *thought* they did it, we were also interested in seeing how people *actually* did this and what elements resulted in high and low levels of social presence. We ran a field study to explore these questions.

For the field study, we selected 56 participants (29 male, 27 female) from an in-house database of usability subjects. 59% of participants were in their 20's, 25% were in their 30's, 9% were in their 40's, none were in their 50's, and 7% were 60 and above. Participants signed up for this study in pairs. Each participant knew their partner previously and was in a different location from them during the study. 13 of the pairs were same-gender pairs (7 male, 6 female), while the remaining 15 pairs were mixed-gender. 36% of pairs were in romantic relationships, 32% were close friends, 29% were immediate family and 7% were acquaintances. 96% of participants reported using video chat applications (e.g., Skype, FaceTime, Google+ Hangouts) at least once or twice a year – 27% using them on a weekly basis.

Participants scheduled a time with their remote companion to watch a video program (e.g., broadcast TV, DVD, internet streaming) together. 15 minutes prior to watching the program, participants logged into Skype and initiated a group video chat with their remote companion. For an additional gratuity credit, participants had the option of adding our auto-answer Skype account to the group video call so we could record it. Once the group video chat was established, participants started the video program at the same time. While the program was running, participants were encouraged to act as they normally would while watching video programs together. After the video program was complete, participants disconnected the video call and completed an online survey. Participants were rewarded for their participation with credit that could be traded for company merchandise. Figure 1 shows a typical set up pairs used to complete this study.
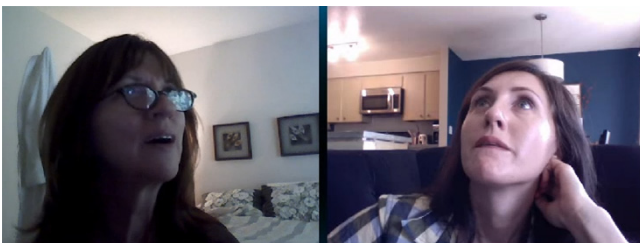

Fig. 1. A field study pair; both watched a program on TV and used video chat on their laptops

We studied the quality of their shared experience, the role of the communication device and the issues and events that affected the quality of their experience. To measure the quality of their shared experience, we looked at social presence, enjoyment and closeness. We used Biocca et al.'s definition of social presence, described as "a sense of co-presence with a mediated person and an awareness for their psychological, emotional and intentional state" [3]. We selected nine questions from the Networked Minds Social Presence Scale [4] most relevant to the nature of our study. Each question was asked on a 7-point Likert scale[1]. We also asked participants to rate how close they felt with their remote companion after completing the activity on a 7-point Likert scale. This construct was added since Kirk et al.'s work described closeness as the main reason behind people's usage of domestic video chat [20]. Finally, we asked them to rate how much they enjoyed the activity on a 7-point Likert scale.

We explored the role of the communication device in various ways. First, we investigated if the communication device was different or the same as the device streaming video programs. Second, we looked at the placement of the device. Last, we measured how the setup affected the video program watching experience in terms of connectedness and interference. We measured connectedness by asking participants to rate how much video chat made them feel connected to their remote companion during the activity. We measured interference by asking participants to rate how much video chat interfered with their experience of watching TV. These measures were in the post-task survey but also examined in the video data as well.

To understand the issues that affected each pair's shared experience, we analyzed it from a technical and context-specific perspective. Technically, we collected audio and video quality ratings of the video chat on 7-point scale plus a rationale behind each rating via open-ended responses within the post-task survey. We also looked at the video data for any technical issues that arose. Contextually, we qualitatively analyzed the video data based on the relationship between a pair of participants, where their attention was focused, and the type of interaction that occurred between each pair.

3.3   Lab Study

The do-it-yourself nature of the field study and its between-subjects design introduced significant variability that made it difficult to compare the different device configurations. We ran a within-subjects lab study that focused on comparing the differences between watching video programs in the same room, watching video programs and using video chat on the same device, and watching video programs and using video chat on different devices.

Eight pairs of participants were recruited from the in-house usability database. They were screened for watching video programs and using video chat on a fairly regular basis, and enjoying video programs with others. All were from the Puget Sound area (6 male, 10 female) ranging from 20 to 60 years of age (median 31).

After being briefed on the study and giving their consent, participants chose a 60 minute television program that they would be interested in watching during the study.

---

[1] All 7-point scales are reported with 1 = most negative, 4 = neutral, and 7 = most positive.

We had five prerecorded shows, which included drama, comedy, and a sporting event. Pairs tried three conditions of watching together – two being remotely connected using video chat. After each condition, they individually completed a short survey; because all questions were optional the number of responses to any given question varied. At the end of the study, they participated in a post-task interview. The study took approximately two hours to complete and participants were rewarded for their participation with credit that could be exchanged for company merchandise.

There were three experimental conditions. Our baseline condition, *Local*, had participants watching a third of the program side by side (Figure 2A). In the *Picture-in-Picture (PIP)* condition they watched a third of the program in separate rooms. The remote person's video was overlaid in the corner of the TV screen and their audio came from a speaker placed near the TV (Figure 2B). Our final condition, *Proxy*, was like PIP but the remote person's video and audio came from a separate device placed near them (Figure 2C). Condition ordering was counterbalanced. In the remote conditions we used high-quality speakers, high-definition webcams, and lavalier microphones to mitigate audio crosstalk.

Fig. 2. Lab study conditions: (A) Local – watching in the same room; (B) Picture-in-Picture (PIP) – watching with remote person appearing in inset on TV, which is highlighted here with an orange rectangle; (C) Proxy – watching with remote person on separate device

We reused our *shared experience* measures from the field study (social presence, enjoyment and closeness) and integrated them into our post-condition surveys. We did this to compare the experiences in the field and lab studies for any overlapping trends. In addition to the quality of shared experience, we expanded on the concept of *enjoyment* and revisited *connectedness* by asking participants to state which condition they enjoyed most/least as well as which condition they felt most connected with their companion. We asked participants to describe the differences between watching video programs remotely and locally. We specifically asked them to reflect on how the second screen's location in the Proxy condition affected their experience. We also asked if they knew anyone with whom they would like to do this activity with in the future.

4   Results and Discussion

Together, our three studies demonstrate that remotely watching TV together is a desirable, fun activity that augments social communication and reinforces social connection. We found that having the communication channels on a separate device was

associated with better outcomes than picture-in-picture on the TV. We found that communication media richness corresponded to better outcomes. Finally we found that there are many technical challenges that can substantially degrade the experience. We present the results of the three studies together to focus on the insights supported by these converging lines of evidence.

4.1   Emerging Practice

Watching video programs, such as television, movies and online content, is often a social activity: 30% of respondents to our initial survey reported watching video programming with one or more other people every day; 75% reported doing it daily or weekly (n=105). At the same time, a substantial amount of watching is done solo: 40% reported watching video programs alone every day; 72% reported doing it daily or weekly (n=95, Figure 3).
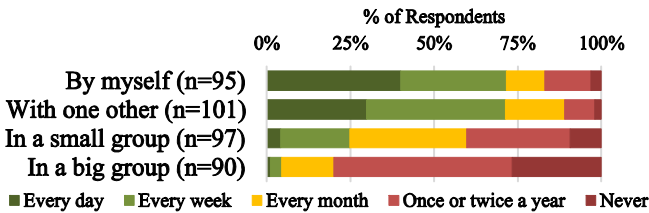
Fig. 3. Responses to the initial survey question, *"When watching video media, I watch it..."*

Some of our respondents (24%) had already communicated with someone in a different location while watching the same video program at least once, using a variety of communication channels, while others (26%) expressed interest in trying it (n=105). Significantly more young people (<30 years old) reported having watched video programs together remotely than older people (39%, n=49 vs. 11%, n=56, Mann-Whitney U=987, *p*=.001, Figure 4).
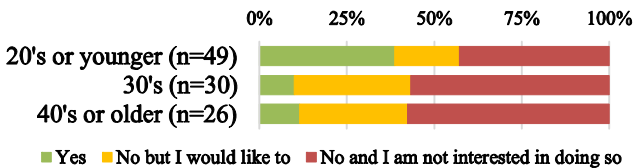
Fig. 4. Responses to the initial survey question, *"Have you ever communicated with someone at a different location while watching the same video program?"* by age bracket

While many of our respondents wanted to watch together remotely, 50% of our initial survey respondents reported being disinterested in communicating with someone at a different location while watching the same video program (n=105). When exploring reasons behind this disinterest, two consistent themes emerged. Some people are simply not interested in watching video programs:

"TV isn't a primary focus of my activities." [Survey P7][2]

Others reported that they do not like communicating verbally while watching:

"In general I don't find watching TV together as a greatest social activity. People either watch and don't speak with each other, or speak with each other but then don't watch. With that I don't get the point of doing it together online…" [Survey P39]

Prior research also found that some participants find watching video programs as time alone to unwind from work [15]. Other participants found certain genres like news and films to require more conscious attention on the video program and allowed less attention for socializing [11].

Following our studies, most participants responded positively when asked if they would be interested in watching video programs with a remote partner again (field study: 88%, n=56; lab study: 88%, n=16). This interest is substantially higher than the combination of the 24% who had tried it and 26% interested in trying it in our initial survey. This difference suggests that getting people over the hump of first trying it may increase the potential audience for this kind of remote shared activity. Harboe *et al.*'s found similar results with sharing audio chat while watching TV, although our results demonstrate video chat as a desirable form of SocialTV [15].

> **Insight:** Some people, particularly young adults, are creating their own remote video watching experiences. There is an opportunity for bringing this capability to a larger audience, especially once they have had the opportunity to try it.

4.2   Enjoyment

Most participants (80%) in our field study responded positively when asked if they enjoyed the experience (n=55, M=5.4, SD=1.5). Several field study participants also commented on their enjoyment using Skype while watching together:

"It is fun because you can comment and talk about things that are happening. [My study partner] and I watched a lot of tv together in college and it really reminded me of that experience." [Field P15a]

Likewise, most participants in our lab study had enjoyable experiences in the remote conditions with 81% responding positively (i.e. *somewhat agree*, *agree*, or *strongly agree*) in enjoyment (n=31, M=5.6, SD=1.3). A significant main effect of condition (Local, PIP and Proxy) was found on enjoyment (Friedman, X^2=17.92, p<.001). Posthoc pairwise comparisons revealed no significant difference between the Local and Proxy condition (Z=-1.996, p=.046) but the PIP condition was rated significantly lower than both the Local and the Proxy conditions (Z=-3.204, p=.001 and

---

[2] Participants are identified by study phase (Survey, Field or Lab), participant number, and a designator (a or b) if they were part of a pair.

Z=-2.705, p=.007 respectively). During the post-study interview, some participants described watching video programs remotely as a fun way to spend time together:

> *"Sometimes you can't be together but want to have Survivor parties or something. It's a great way to hang out if you actually couldn't."* [Lab P7a]

A few participants reported that using Skype while watching the same video program was initially awkward:

> *"It was a new experience and slightly awkward at first but then realizing we were laughing at the same things was fun"* [Field P17b]

Several participants reported, unsurprisingly, that it was not as pleasant as watching together in person.

> *"I still was able to easily communicate with my companion, as well as see their instant reactions to exactly what I was watching. However, because I was not physically in the room with them, it didn't seem to be as enjoyable."* [Field P13b]

Previous research provides possible reasons for this preference. Gaze and gestures are difficult to interpret during video chat, making an assessment of body language difficult [17]. Participants in video chat have a harder time reading their companion's mood and attention which can hinder effective interaction and negatively affect the shared experience [14]. Similarly, video chat lacks a shared physical space. Some people simply feel that digital representations of their companion are insufficient in comparison to having them physically there [16]. While we did not expect it to be as good as watching together in the same room, our goal was to explore enabling a shared watching video experience when they could not be together physically.

Some participants felt that the communication enhanced their enjoyment of the content.

> *"It was far more enjoyable than watching it alone; it's nice to discuss the show or make comments!"* [Field P27a]

Others felt that the communication lessened their enjoyment of the content.

> *"I'm more distracted from the video program and can't fully enjoy it."* [Field P17a]

The field study questionnaire results also showed mixed results regarding whether communication interfered with the content: 31% agreed, 40% neutral, and 29% disagreed (n=55, M=4.0, SD=1.6).

Insight: Most people enjoy the experience of remotely watching video, but challenges remain to make it more natural and to reduce the conflict between the video program and the communication.

4.3   Social Presence and Connectedness

We used nine questions from Biocca & Harms' Social Presence Inventory [4], along with our additional question of *closeness*, to gauge participants' level of social presence in the field study (Figure 5).

SP1. I often felt as if my remote companion(s) and I were in the same room together.

SP2. My remote companion(s) were often aware of me.

SP3. My remote companion(s) paid close attention to me.

SP4. I was easily distracted from my remote companion(s) when other things were going on. (Note: Response scale was inverted for analysis and presentation.)

SP5. I was sometimes influenced by my remote companion(s)' moods.

SP6. My remote companion(s) were able to communicate their intentions to me.

SP7. My thoughts were clear to my remote companion(s).

SP8. My actions were often dependent on my remote companion(s)' actions.

SP9. The behavior of my remote companion(s) was often in direct response to my behavior.

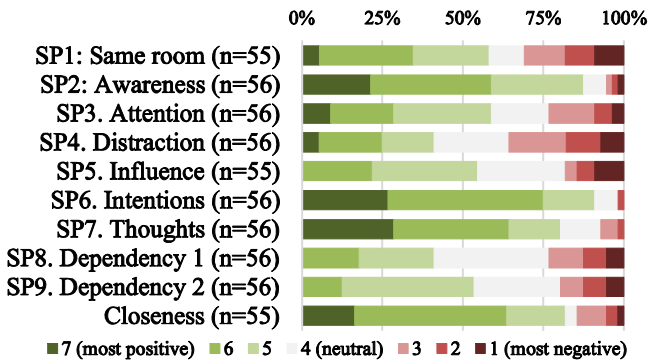- -   Watching video media together remotely made me feel closer to my remote companion(s).

Fig. 5. Responses to the field study post-test survey questions relating to social presence

In order to compute a composite social presence score, we utilized a principal components factor analysis to extract relevant factors from these ten questions. Two different factors (factor load greater than .5) were revealed. The first factor included six of the questions from the Social Presence Inventory (SP1, SP2, SP3, SP4, SP6, and SP7) along with the closeness question. The second factor included the remaining three questions from the Social Presence Inventory (SP5, SP8, and SP9).

The first factor was significantly ($p<.01$) correlated with users' feelings of connectedness ($r=.75$), their enjoyment ($r=.60$), and their willingness to do the activity again ($r=.52$). As a result, we used the mean of these seven measures to compute the composite "Social Presence" (*SP*) score (the resulting score was in the range of 1-7, with 1 being low and 7 being high). The second factor was not significantly correlated with any of the other measures ($r<.1$) so these questions were not used in subsequent analyses.

In the field study, 86% of participants' SP scores were positive (>4) (M=5.09, SD=.97). Examining all of the demographic data for the participants (gender, age, level of experience with Skype, relationship with their partner), as well as their equipment configuration (device used for program and communication, and placement of the device) revealed no significant impact on the resulting SP scores (using Kruskal-Wallis test with alpha=.01). This may be a result of high individual differences across these variables, as well as the fact that most participants enjoyed their experience, regardless of the setup.

We used the same measure of SP in the lab study (Figure 6). For the two remote conditions, participants' enjoyment ratings were significantly correlated with their SP scores (PIP: r=.74, p<.001; and Proxy: r=.54, p=.027). This correlation supports previous work that found individual satisfaction (or enjoyment) from a leisure activity was the most influential factor on relationship satisfaction [2]. A significant main effect of condition (Local, PIP and Proxy) was found on SP scores (Friedman, X^2=30.63, p<.001). Posthoc pairwise comparisons revealed that Local had significantly higher SP scores, than Proxy (Z=-3.367, p=.001) or PIP (Z=-3.724, p=.001), and Proxy had significantly higher SP scores than PIP (Z=-2.976, p=.003). These results support prior literature comparing VMC to face-to-face communication [16,17]. They are also consistent with Media Richness Theory [10], suggesting that the communication media fidelity plays a strong role in the social connection of the experience.
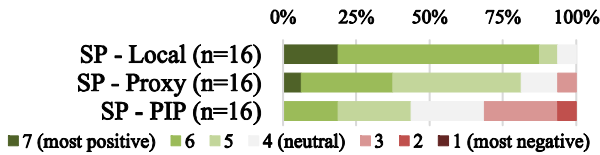

Fig. 6. Social Presence scores from the lab study post-condition survey. (Note: SP scores have been rounded to nearest integer.)

Most participants expressed that using VMC while watching video programs made them feel connected with their remote partner. In the field study, 75% of the participants responded positively that the experience made them feel connected (n=56, M=5.3, SD=1.3). Qualitative data from the field study provides some reasons why:

> *"Being able to hear comments from my friend helped me feel connected with him."* [Field P4a]

> *"It made me feel like I was sharing the same emotion with him...and we laughed etc together!"* [Field P8b]

> *"My sister lives two states away, it felt as if we were sitting next to each other watching a show like the old times."* [Field P16a]

Other comments suggested that the reason for feeling connected could be the common ground provided by the shared activity.

> *"You get to enjoy something together, rather than just talk."* [Field P2b]

> *"just skyping makes me feel as i need to be talking about something when i dont really have something to say but when watching media we can comment on whats going on"* [Field P12b]

> *"Well, when you are doing the same thing while Skyping, you have that little bit more in common with them. It definitely makes it more satisfying."* [Field P24a]

These results matched our hypothesis that the video program could serve as a shared space and coincides with Clarke and Brennan's concept of grounding [8]. The shared reference enabled participants to spend more time enjoying the activity and discussing the show or personal topics rather than spending the time trying to establish a shared perspective.

On the other hand, there is some design tension around competing video channels as some participants reported that watching the video programs distracted them from their normal Skype conversation behavior and decreased their sense of connectedness.

> *"When I usually Skype, I'm not doing anything else but Skyping, so it was a little strange because I felt like I was ignoring her... I think it takes away from the chat because you are watching the video therefore neither one gets your full attention"* [Field P11a]

> *"when I Skype with my husband it is usually to show him the baby or dog or important mail. Not eat up our time with a tv show"* [Field P25b]

O'Hara *et al.*'s work on mobile video chat [25] observed similar findings. Their participants described VMC as mentally demanding and, unlike mobile audio chat, made multitasking more difficult because the other party could see that you were not fully attending to them.

Insight: For most people, remotely watching video resulted in a strong sense of connection or social presence. It strengthens ties when framed as an additional activity that increases the time spent together rather, than taking time away from talking together; controlling this framing is a significant design tension.

4.4   Placement of the Skype Window

In the initial survey, 48% of respondents ranked the picture-in-picture configuration most highly, where the video of their remote partner is shown embedded on the display showing the video program. 24% respondents wanted it on a separate device close to them, 16% wanted it on a separate device near the television, and 12% wanted it on a device they were holding (n=25).

In the field study, where users were able to choose whichever configuration they preferred or had access to, 61% used a picture-in-picture configuration and 39% used a Proxy configuration (n=54). Mann-Whitney U tests did not detect any significant differences between PIP and Proxy configurations in terms of enjoyment ($Z=-.853$, $p=.394$), SP ($Z=-.116$, $p=.908$) or their desire to do the activity again ($Z=-.055$,

p=.956). The lab study however showed a clear benefit of Proxy over PIP (Wilcoxon Signed Ranks Test) for both participants' level of enjoyment (Z=-2.705, *p*=.007), and SP scores (Z=-2.976, *p*=.003).

In the lab study post-test interview, eight participants reported liking the proxy in the position we had set, five wanted it at the same distance but more in line of sight of the television, and three wanted it farther from them and closer to the TV. Note that the Proxy could be positioned to more easily see both the video content and the remote person than the head turning needed if sitting side-by-side in the same room.

It should be noted that the PIP conditions of the initial survey and lab study are not identical to that of the field study. In the former, PIP referred to both content and communication being placed on a television screen viewed from a distance. In the latter, all PIP configurations devised by our participants were on a laptop or desktop viewed close up.

Media Richness Theory [9] suggests that the Proxy condition would be preferred over PIP. The Proxy condition presented the remote companion on a larger screen, allowing for a higher quality representation of the individual. Likewise, the separation of TV and Proxy speakers allowed the participant to directly perceive the sound sources as spatially distinct. The visual and audio representations came closer to emulating the remote individual's physical representation and as a result made participants feel closer to having the actual person in the room [1,10].

Likewise, Grounding Theory [8] can further explain the Proxy's success over PIP. In the Proxy condition, a participant receives a visual cue when the other person shifts their visual attention between the content and the remote view. Because of this, less effort was needed to reorient between focused conversation and attending to the video program. For example, as seen in the video data, a participant could tell if their remote companion's attention was primarily focused on them based on how often they would turn their gaze toward the video chat screen. The participant in return could reciprocate and look back at the remote companion or keep looking at the TV, indicating that they are engrossed in the show and do not wish to chat right now.

> **Insight:** When the video program is on a television, having the remote person on a proxy device on the coffee table is much preferred over picture-in-picture. Future work should compare these conditions to picture-in-picture on a laptop or tablet, as our results were inconclusive.

4.5   Communication Channels

Remote social TV can employ a variety of communication modalities. Prior work has focused on textual [7,9,11,24,26] or audio [12,15] modalities. The use of VMC in this application is largely an unexplored design space. Respondents to our initial survey ranked which modalities they felt were most important to help them feel connected. Talking verbally (i.e. audio) was rated highest, followed by text, then visual, then touch (Figure 7).
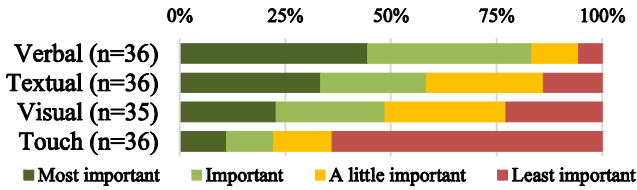
Fig. 7. Responses to the initial survey question, *"What interaction is needed to feel most connected with the person you are watching TV remotely with?"*

As noted above (Figure 6), in the lab study the Proxy condition had more positive outcomes than PIP. The most notable difference between these conditions is the greater visual angle subtended by the remote video. This result suggests that the video of the remote partner adds a key element to the experience.

While Figure 7 shows that touch was the lowest-ranked medium, 10% of respondents ranked it #1 and 10% ranked it #2 (n=36). In the field study questionnaire we asked, *"If there was a magical device that could increase your feeling of being connected with your companion(s), what key function or property does it need to have?"* Several respondents described wanting some form of touch:

> *"For me it's spatial. I want to feel warmth from loved ones in close proximity... so something that could make me feel like she was sitting right next to me."* [Field P16b]

Many researchers have investigated haptic interpersonal communication (e.g. [5]) and watching video together remotely may be a scenario that is ripe for exploring it.

> **Insight:** While audio is fundamental for communication, video appears to play a key role in creating social presence. Tactile communication deserves more study.

4.6   Technical Problems

There were many technical hurdles that our field study participants needed to surmount, including synchronizing the content, handling the multiple audio streams, and general computer/networking glitches. Pairs spent a median of 3.6 minutes on setup (n=24). Twelve pairs continued to have technical problems throughout the session. We discarded the data of one pair because they were unable to solve the technical issues. 28% rated their audio experience negatively (n=54, M=4.6, SD=1.5); and 22% rated video negatively (n=55, M=5.1, SD=1.5).

These ratings were significantly correlated with participants' enjoyment (audio: r=.40, p=.003; video: r=.36, p=.008), SP (audio: r=.454, p=.001; video: r=.315, p=.019), and willingness to do the activity again (audio: r=.356, p=.008; video: r=.275, p=.042). This supports previous work that lists technical components as one of three factors (along with social and psychological) that affect the experience of communication [1].

While Liu *et al.* have developed a system for synchronizing video in a popular IM system [23], there is currently no widely-available standalone mechanism for

perfectly synchronizing video playback between a pair of computers in different locations. For prerecorded programming, most participants used a technique where one counted down and then both pressed *Play*. Even broadcast programming is not perfectly synchronized since different broadcasting pipelines (e.g., satellite, cable, antenna) introduce different delays. While sync issues can be less noticeable in text chat, the slightest sync issue is much more obvious in video chat. Eight pairs in the field study had sync issues. Four pairs mitigated the sync issues by muting the content audio at one side and listened to it via Skype audio. When the content was out of sync even slightly, it impeded the interaction between the participants.

> *"My companion became annoyed that I was trying to make it so we were watching the shows at the exact same time. But if we weren't it was strange because we would laugh in different places."* [Field P18b]

These findings replicate similar findings from studies that suggest minor sync differences within video chat have significant negative effects on communication [21,22].

Audio was problematic because the microphone, which was intended to pick up the viewers' speech, also picks up the video program's soundtrack. While the automatic echo cancellation capabilities of the computer can keep the speech signal clean, it typically could not suppress the echo of the video program audio. When this occurred, a participant would hear the program audio both directly and via the communication channel, adversely affecting the engagement:

> *"Imagine sitting in a room watching TV, and having another TV in the same room, same show, but the timing is off a little bit and the volume is turned up."* [Field P19b]

Even without audio crosstalk, the two audio streams competed. Speech could drown out the program content, or vice versa.

> *"it was hard to hear the other person when they laughed or what not."* [Field P25b]

Insight: Technical complexity, content synchronization, and audio crosstalk are barriers to remote video watching. Technical problems can decrease the effectiveness of the overall experience.

---

5   Conclusion and Future Work

We found that remotely watching video programs together while connected by VMC is engaging, fun, and fosters social bonds between the participants. Our study documents the important factors that enable this particular "shared experience" [6] to lift VMC "beyond talking heads."

We found little to suggest that the combination of audiovisual entertainment and audiovisual communication caused divided attention. Audio levels did compete, but having two separate video streams did not. Duchenaut et al. [12] noted that the pauses

and redundant content in TV shows create opportunities for communication and may mitigate this competition. The Proxy configuration's advantage over PIP may have been in part because it helps people better manage their attention. We believe that there is further opportunity for better understanding and designing for attention in shared-experience systems.

We believe that a key affordance of VMC is how it enabled more immediate and natural sharing of reactions, which is an important part of enjoying watching videos together. Video, compared to text or only audio, is well-suited for richly sharing laughter and other emotional reactions to the video program. The users' comments suggested that this sense of sharing these reactions through video contributed to the positive scores in social presence, connectedness, and enjoyment.

We believe that practical systems for remotely watching video together would have positive social impact by allowing close ties to strengthen their bonds by sharing experiences despite physical distance in a way that is different from, and complementary to, simply conversing together. Our next step is to develop software – and possibly hardware – to support watching together remotely. There are many interesting design and technical problems – initiating the experience, choosing the program to watch, closely synchronizing playback, and solving audio crosstalk.

Finally, watching TV is but one of many possible remote shared experiences. This study strongly supports rich media beyond audio communication in remote shared experiences. This is a rich design space that deserves more exploration.

References

1. Aaltonen, V., Takatalo, J., Hakkinen, J., Lehtonen, M., Nyman, G., Schrader, M.: Measuring Mediated Communication Experience. In: QoMEx 2009, pp. 104–109. IEEE (2009)
2. Berg, E.C., Trost, M., Schneider, I.E., Allison, M.T.: Dyadic Exploration of the Relationship of Leisure Satisfaction, Leisure Time, and Gender to Relationship Satisfaction. Leisure Sciences: An Interdisciplinary Journal 23(1), 35–46 (2001)
3. Biocca, F., Harris, C., Burgon, J.: Toward a more robust theory and measure of social presence: review and suggested criteria. Presence: Teleoperators and Virtual Environments 12, 456–480 (2003)
4. Biocca, F., Harris, C.: Networked Minds Social Presence Inventory (Scales only, ver. 1.2) (2003), `http://cogprints.org/6742/1/2002_netminds_scales.pdf` (accessed September 19, 2012)
5. Brave, S., Ishii, H., Dahley, A.: Tangible interfaces for remote collaboration and communication. In: CSCW 1998, pp. 169–178. ACM Press (1998)
6. Brubaker, J., Venolia, G., Tang, J.: Focusing on Shared Experiences: Moving beyond the camera in video communication. In: DIS 2012, pp. 96–105. ACM Press (2012)
7. Cesar, P., Chorianopolous, K., Jensen, J.J.: Social Television and User Interaction. ACM Comput. Entertain. 6(1) (2008)
8. Clark, H.H., Brennan, S.E.: Grounding in communication. In: Resnick, L.B., Levine, R.M., Teasley, S.D. (eds.) Perspectives on Socially Shared Cognition, pp. 127–149. APA, Washington, DC (1991)
9. Coppens, T., Trappeniers, L., Godon, M.: AmigoTV: Towards a social TV experience. In: Proc. EuroTV 2004, vol. 36 (2004)

10. Daft, R.L., Lengel, R.H.: Organizational information requirements, media richness and structural design. Management Science 32, 554–571 (1986)
11. Dezfuli, N., Khalilbeigi, M., Mühlhäuser, M., Geerts, D.: A Study on Interpersonal Relationships for Social Interactive Television. In: Proc. EuroITV 2011, pp. 21–24. ACM Press (2011)
12. Ducheneaut, N., Moore, R.J., Oehlberg, L., Thornton, J.D., Nickell, E.: Social TV: Designing for distributed, sociable television viewing. International Journal of Human-Computer Interaction 24(2), 136–154 (2008)
13. Fitzpatrick, T.R.: The Quality of Dyadic Relationships, Leisure Activites and Health Among Older Women. Health Care for Women International 30(12), 1073–1092 (2009)
14. Flora, J., Segrin, C.: Joint Leisure Time in Friend and Romantic Relationships: The Role of Activity Type, Social Skills and Positivity. Journal of Social and Personal Relationships 15(5), 711–718 (1998)
15. Harboe, G., Massey, N., Metcalf, C., Wheatley, D., Romano, G.: The Uses of Social Television. Computers in Entertainment 6(1), article 8 (2008)
16. Heath, C., Luff, P.: Disembodied conduct: Communication through video in a multimedia environment. In: CHI 1991, pp. 99–103. ACM Press (1991)
17. Hirsh, S., Sellen, A., Brokopp, N.: Why HP People Do and Don't Use Videoconferencing Systems (2005)
18. Inkpen, K., Taylor, B., Tang, J., Junozovic, S., Venolia, G.: Experiences2Go: Sharing Kid's Activities Outside the Home with Remote Family Members. In: CSCW 2013. ACM Press (2013)
19. Junozovic, S., Inkpen Quinn, K., Blank, T., Gupta, A.: IllumiShare: Sharing any surface. In: CHI 2012, pp. 1919–1928. ACM Press (2012)
20. Kirk, D., Sellen, A., Cao, X.: Home Video Communication: Mediating 'Closeness'. In: CSCW 2010, pp. 135–144. ACM Press (2010)
21. Krauss, R., Bricker, P.D.: Effects of transmission delay on the efficiency of verbal communication. Journal of Acoustical Society of America 41(2), 286–292 (1967)
22. Kraut, R.E., Gergle, D., Fussell, S.R.: The Use of Visual Information in Shared Visual Spaces: Informing the development of virtual co-presence. In: CSCW 2002. ACM Press (2002)
23. Liu, Y., Shafton, P., Shamma, D.A., Yang, J.: Zync: the design of synchronized video sharing. In: Proc. DUX 2007, article no. 12. ACM Press (2007)
24. Metcalf, C., Harboe, G., Tullio, J., et al.: Examining presence and lightweight messaging in a social television experience. ACM Trans. Multimed. Comput. Comm. Appli. 4(4) (2008)
25. O'Hara, K., Black, A., Lipson, M.: Everyday practices with mobile video telephony. In: CHI 2006, pp. 871–880. ACM Press (2006)
26. Shirazi, A.S., Robs, M., Schleicher, R., Kratz, S., Müller, A., Schmidt, A.: Real-time nonverbal opinion sharing through mobile phones during sports events. In: CHI 2011, pp. 307–310. ACM Press (2011)
27. Whittaker, S.: Things to talk about when talking about things. Human-Computer Interaction 18, 149–170 (2003)