

# Social TV: Designing for Distributed, Sociable Television Viewing

Nicolas Ducheneaut<sup>1</sup>, Robert J. Moore<sup>1</sup>, Lora Oehlberg<sup>2</sup>, James D. Thornton<sup>1</sup>, Eric Nickell<sup>1</sup>

<sup>1</sup>Palo Alto Research Center; <sup>2</sup>UC Berkeley, Mechanical Engineering

[nicolas, jthornton, bobmoore, nickell]@parc.com

## Abstract

Media research has shown that people enjoy watching television as a part of socializing in groups. However, many constraints in daily life limit the opportunities for doing so. The Social TV project builds on the increasing integration of television and computer technology to support sociable, computer-mediated group viewing experiences. In this paper, we describe the initial results from a series of studies illustrating how people interact in front of a television set. Based on these results, we propose guidelines as well as specific features to inform the design of future “social television” prototypes.

## 1. Introduction

Television has often been criticized as an isolating, anti-social experience. However, early ethnographic studies (Lull, 1990) showed that

*“TV and other mass media, rarely mentioned as vital forces in the construction or maintenance of interpersonal relations, can now be seen to play central roles in the methods which families and other social units employ to interact normatively.”*

And indeed, television viewing appears to be largely a social activity, often conducted in groups (Morrison, 2001). In fact, the worth of a particular television program is often gauged according to the amount of social interaction it generates (White, 1986). Television can foster multiple forms of sociability: *direct* (e.g. when chatting with friends and family during a “movie night” at home) or *indirect* (e.g. when discussing previously viewed programs with colleagues at the office water cooler), and both are equally worthy of attention. Previous research (Morrison, 2001) highlighted a similar distinction between the “internal” social functions of television viewing (when family members watch television together) and its “external” functions (e.g. television programs as topics of conversation at work or elsewhere; special events organized at home such as inviting friends over for watching the Superbowl).

While there is now little doubt watching television can be a “ticket to talk” (Sacks, 1992) encouraging interaction between groups of viewers, there is little research available on the exact practices surrounding sociable television viewing. Media research has focused essentially on the *role* that television plays in social groups like the family, examining for instance how programs are interpreted outside the viewing context and how this contributes (or not) to the maintenance of the family system (Alexander, 1990). This research tends to be based on surveys and ethnographic observations of group viewing like Lull’s (1990) remain rare. Audiences have been generally considered as “a category rather than a way of being” (Casey et al., 2002) – that is, past research has investigated how a given audience’s socio-demographic characteristics affect their choice of programs, rather than how they behave as active participants engaged in social activities while watching television.

As such, the actual *mechanics* of joint viewing (e.g. When do people talk during a show? About what? To what effect?) remain largely unexplored (an interesting exception can be found in Walker and Bellamy, 2001, a survey of remote control use during family viewing). This lack of data has important design implications: a better knowledge of joint viewing practices could help develop new technology to better support television-mediated sociability.

It might seem that practices around a technology as well-established as television are not amenable to redesign, but in fact both societal change and technological innovation have been affecting viewing, as well as entertainment in general, for a long time. Many characteristics of contemporary life in wealthy societies conspire against traditional joint television viewing, despite its social appeal. Urban sprawl, for instance, can make travelling to a friend’s house or a public space for a movie night inconvenient (Oldenburg, 1989); domestic isolation and scheduling constraints prevent gatherings (Putnam, 2000); and increasing mobility often separates family members (e.g. a child living away from his family to attend university). Sociability is becoming more and more distributed in this context

as technology enables diverse remote interactions. Online computer games, for instance, are displacing television among young viewers, in part because they offer an interactive, social, and location-independent experience (Subrahmanyam et al., 2002; Schwartz, 2004).

Technology is changing television practices also, through integration of computing capabilities as in the case of digital video recorders such as TiVo<sup>1</sup>. Our Social TV project is about leveraging this kind of computing integration to remove the increasing barriers to sociable interaction around video content. In focusing on the direct form of sociable viewing we will be talking about design for *distributed, shared television viewing*. While this necessarily changes the television experience, our concern is to preserve the 'natural', familiar social atmosphere of watching television in a collocated group. This goal of enhancing social opportunity while fitting existing practice frames the central design challenge.

In this paper, we report on preliminary studies and design concepts that serve our goal. We begin with observation of what television viewers do while they are watching television in groups, both when they are physically co-present and geographically distributed. Based on our observations, we then articulate in the second part of this paper a series of design guidelines and concepts for future Social TV prototypes.

## 2. Related Work

Surprisingly, very few systems have been developed to support social interactions of any kind among television viewers. Two projects are closest to ours. The first one is Chuah's (2002) "reality instant messenger", which provides both "buddy surfing" (an awareness that friends are watching the same television program) and an IM-based communication channel between viewers. The second related project is Alcatel's Amigo TV (Coppens et al., 2005), which allows television viewers to share opinions and feelings with friends via an interactive broadband link. Amigo TV shares a similar design philosophy to ours but, as we shall see later in this paper, the range of features offered to users is quite different.

Another approach has been to transform television into an inhabited virtual world. In Benford et al. (1998), audience members control avatars in a 3D space and can interact with the performers of the show they are watching. The focus is on breaking

down the barriers between audience members and performers, as opposed to facilitating group interaction while watching television.

## 3. Understanding TV-Mediated Sociability

In order to understand what can make group television viewing so enjoyable, we invited participants to watch various television shows together in a specially equipped room in our laboratory. Our experiments were meant to reproduce one prevalent form of group television viewing: the "viewing party" where, for instance, a group of friends get together to watch the latest episode of the show "Desperate Housewives." It is important to recognize, however, that other forms of group television viewing are equally interesting and that they would probably affect viewing patterns differently. For instance, we plan to investigate "everyday viewing" (e.g. married couples watching television together in their own home, without planning) at a later date.

We began with three exploratory sessions where viewers were all located in the same room and could interact face-to-face. The smallest group size was 5 and the largest 8. Participants were between 20 and 50 years old, with more males than females (70% and 30% respectively). Most were colleagues but a few shared social activities outside of work. Participants watched a soccer game or one of two episodes of a documentary (the BBC's "1900 House"), each lasting approximately 2 hours overall. We used two cameras to videotape both the participants' behaviour and the content of the television show. The tapes were later reviewed and coded (Glaser, 1998) by two of the authors, with a particular attention to the local organization of talk among the participants (Sacks et al., 1974). Our observations were complemented by a questionnaire asking participants about their group television viewing habits (e.g. how often they viewed TV in groups, the kinds of content watched, etc.), as well as exit-interviews after the viewing session to reflect on their experience.

In the second phase of our study we held six additional viewing sessions, this time separating the participants into two groups located in two different rooms. A social audio link was established between the rooms using two computers running Robust Audio Tool (RAT<sup>2</sup>, see Figure 1). By relaying all audio between the two rooms, each group could hear what the other was saying at all times. While the audio quality was adequate, it was certainly degraded compared to co-present interactions.

---

<sup>1</sup> <http://www.tivo.com>

---

<sup>2</sup> <http://www.netsys.com/mbone/software/rat/>

However, the quality was good enough to allow our experiments to proceed.



**Figure 1 – Equipment used during Phase 2**

These later experiments were meant to simulate some of the conditions users might encounter while watching television in a distributed setting. We chose not to transmit video streams of the participants' faces between the rooms, as this has already been tried elsewhere (Huijnen et al., 2004). Our focus was on the most basic aspect of sociability: conversation.

As before, we used cameras (this time, three of them) to videotape both the participants and the audience. We assembled and synchronized the three video feeds such that they could be reviewed and coded on a single screen (see Figure 2). We administered the same questionnaire after each session as was used in phase 1.



**Figure 2 – Merged video feeds used for coding our participants' interactions**

In the second phase, there were at least 2 and up to 6 participants in each room. The gender ratio was the same as before, and so was the age range. We

invited again some of the participants from the first phase to allow them to compare and contrast the two conditions. Participants watched a wide range of content ranging from sporting events (the Olympics) to cartoons (the Simpsons) and reality-TV shows (the Real World).

### *3.1. Survey Results*

The majority of the survey respondents reported that they routinely held viewing parties with their friends, and among the reasons why were that "TV [is] an excuse for sociality... [we watch] what[ever] TV is "big" enough to justify having someone come over." This confirmed both our intuitions and earlier research about the attractiveness of group television viewing.

We also asked people what kinds of television programs they watched in groups. Among the responses the most popular genres included Animation, Sports Events, Documentaries, Action-Adventure, and Reality Television. It is clear from our participants' comments that certain qualities in TV shows encourage sociability more than others. In particular, shows with bursty rhythms or redundant content (such as sporting events) provide plenty of pauses and opportunities for interaction. People-centered content (such as reality TV shows) provide audiences with many "conversational props" (Lull, 1990). Indeed, as one of our participants humorously mentioned during an exit interview, "It's fun to make fun of people." This parallels earlier research on the social psychology of entertainment. Stone (1981), for instance, noted that sports spectacles "are conversation-pieces, and conversations about them before, during, and after the event bring people together in an emotional rapport" (p. 222). The talk around the event itself may be interpreted as "a loosely structured game [...] a form of nonutilitarian play" (Crabb and Goldstein, 1991, p.367).

Poor quality movies were also often mentioned as a good way to foster social interaction. Our respondents did not only find it fun to talk about poor casting, acting, or effects: a consensus about a lack of important or relevant dialogue in a show seems to provide viewers with as many opportunities to comment as they would like. Shows such as "Mystery Science Theater 3000"<sup>3</sup> have capitalized on this phenomenon. All the characteristics above encourage a kind of "vicarious audience play" (Sutton-Smith, 2001) that is central to sociable television viewing.

<sup>3</sup> <http://www.mst3kinfo.com/mstfaq/basics.html>

As we were collecting survey results, it became clear that content selection would make a difference to the level of social interaction independent of other features of the experimental condition. This was indeed supported by our observations. For instance, the documentary we later used during the first phase of our experiments clearly generated less lively conversations than other genres (although interesting sociable exchanges occurred nonetheless, as we illustrate later in this paper). To make sure our participants had enough material and incentive to communicate, we therefore purposefully selected the most popular genres for most of our experiments.

### 3.2. Empirical Observations

The most striking finding to emerge from our observations was the surprising similarity in the nature and structure of the participants' conversations across the two experimental conditions. We identified clear interaction rules that participants respected when they interacted during a TV show, whether or not they were collocated.

These rules were never openly discussed by the participants. Instead, they seem to be part of a set of ingrained cultural practices dictating proper behaviour when watching television in groups.

#### 3.2.1. Interactional Practices When Watching Television in Groups

Critics of television often point out that the nature of television programs encourages passivity (Casey et al., 2002). This passivity could be explained, at least in part, by the fact that the television set tends to dominate most channels of communication (both audio and visual) and, as such, it might not be conducive to interactive exchanges between audience members. However, our experiments revealed that television viewers are quite adept at communicating with each other during a show. To do so, they rely on a set of interactional practices that allow them to simultaneously socialize with each other around the TV *and* to follow the ongoing program with sufficient attention.

---

```

01 Voice:   in nineteen hundred it must have been (0.5)
02         fantastic to be able to whiz down the roa:d
03         (0.4)
04 Voice:   on your bicycle.=
05 Music:   =[(gets louder)]
06         [(P1 begins to take a drink)]
07         (2.2)
08         (1.2) ((P2 turns toward P1 and P3))
09 P2:      °(      musta been      [   ] °
10 Voice:   [HE::Y I'M FREE
11         (0.2)
12 P3:      °yeah°
13         (0.1)
14 P3:      [°yeah°]
15 P4:      [wha:t?]
16         (0.8)
17 P4:      [what ro:b?]
18 Music:   [ Fade Out ]
19         (0.5)
20 P2:      well it really [mus]t have been ju[st=
21 Video:   [FO ] [FI
22 Music:   [FI
23 P2:      =liber[ating]
24 P4:      [Oh:: ] yeah::
25         (0.2)
26 Voice:   but not everyone [(in the)] home can=
27 P4:      [ yeah:: ]
28 Voice:   =enjoy their freedom...

```

**Transcript 1 – The participants' interactions are carefully timed to exploit gaps in the show**

---

To illustrate this phenomenon, let us consider the transcript above. The excerpt is based on a three-minute long segment from one of our Phase 1 sessions, during which seven co-located participants were viewing “The 1900 House”, a historical documentary. The transcript is coded using the conventions of Conversation Analysis (Sacks et al., 1974), emphasizing the sequential organization of turn-taking in social exchanges. The excerpt features four of the seven participants, identified as P1-P4 (their talk appears in bold). FI and FO indicate fade-in and fade-out in the video stream. At this moment of the show, the main protagonist (coded as “Voice” and in italics) describes her joy at being able to ride a bicycle.

In the transcript, we can see how the viewers are finely attuned to the structure of the show – a show that none of them has watched before – by the way they insert their talk precisely in the gaps in dialogue and transitions between scenes. This is made possible by the fact that such gaps are more or less *projectable*, much like turns-at-talk (Sacks et al. 1974). There are several gaps in the TV dialogue, but they are not all appropriate places to talk. For example, the 0.5 second silence at the end of line 01 occurs at a place at which the narrator’s utterance is hearably incomplete. In contrast, the 0.4 second silence at line 03 is a possible completion point, although the actual completion point occurs at the end of line 04. Now at the end of line 04, there is a cinematic cue that projects an upcoming scene transition: the music gets louder (line 05). P1 orients to this projected transition nonverbally by beginning to take a drink at precisely this point.

However, it is not until after an additional 2.2 seconds of silence (line 07) in the TV dialogue that P2 turns toward P1 and P3 and makes a quiet comment about the show (line 09). Although this would seem an ideal place to talk, it actually overlaps slightly with the final utterance in the scene (line 10) that appears as a surprise. P2 then receives two quiet agreement tokens from P3 (lines 12 and 14), short responses that minimize the length of the conversational sequence. But in overlap with the second, P2 also receives a request for a repeat from P4 (lines 15 and 17) who is across the room and who apparently could not hear P2’s comment. P4 thus expands the sequence. As P2 repeats his comment louder and directs it to P4, the scene fades to black and then fades back in (line 21) right in the middle of his turn. P4 then marks her recognition of the repeat with “Oh:.” in overlap with his turn and produces a minimal agreement token (line 24) just before the dialogue of the next scene begins (line 26). She then repeats the token in

overlap with the TV dialogue (line 27) and the participants then refrain from speaking as the new scene opens. Thus, the participants work hard to fit their conversation into the transition between scenes.

Beyond the kind of exchanges illustrated above, we also observed that lulls in programming are used to help newcomers, who may have arrived late in the program, catch up on what happened and is currently going on in a program. The composition of the audience is often fluid during group television viewing, with participants constantly leaving, arriving, or coming back. Our participants used gaps in the show to progressively bring their co-viewers up to speed.

Moreover, it is also important to note that, even when the participants had access to a remote control allowing them to pause or stop the show at any time it was almost never used. Instead, the participants preferred to let the show continue at its own pace and relied on their aforementioned ability to predict gaps in order to communicate with each other. This confirms earlier research showing that group viewing creates a certain pressure “to just leave the clicker alone” (Walker and Bellamy, 2001, p.83). Participants structure their interactions to avoid using technological control over the show’s pacing and structure.

Although viewers tend to try to fit their talk within gaps in the program, they are not constrained to do so and in fact sometimes choose not to. An example of such overlap is illustrated in the following case (Transcript 2, next page). In this excerpt, from Phase 2 of our study (three viewers in room A and two in room B), the viewers watched *The Real World*, a reality-TV show.

In the first part of this excerpt (line 05), A2 makes a joke about the car that one of the *Real World* roommates is being driven to the airport in by her father. The joke is precisely timed to fit in a gap in the TV conversation, like the conversation in Transcript 1. However, because the joke is not placed near a scene transition, there is no gap for a recipient to respond without overlapping with the program. Thus, in overlap with the next utterance on the TV, B1 responds nonverbally with a smile (line 08) that A2, in the other room, cannot see; A3 responds with a short utterance that overlaps with the TV only minimally (line 09); and B2 responds with a longer joke (suggesting the backwardness of Alabama) that occurs entirely in overlap with the TV (line 11). Thus, without a ready gap for a response, the participants nonetheless choose to interrupt the TV dialogue to respond to A2.

---

01 TV: =an if you jus' stay in your little neighborhood there  
02 an it looks tuh me like everythang would be:  
03 (0.5)  
04 work out gr[eat.  
05 A2: [they still make cars like that?  
06 (0.1)  
07 TV: {what ti:me} do you think [would be good] ti:mes=  
08 {B1 smiles } [ ]  
09 A3: [ not anymore ]  
10 TV: =f[or you tuh [ca:ll ]me eleven uh] clock?  
11 B2: [ in al[abama?] yeah. ]  
12 A1: [ hhuh ]  
13 (0.3) ((B2 smiles))  
14 TV: .hh (0.1) i- (0.3) in the mornin' an eleven o'clock at night?  
15 (1.2) ((scene transition))  
16 (0.9) ((TV: voice over PA system))  
17 TV: tell me bah!  
18 (0.9)  
19 TV: °hm not like ol[[ ]  
20 A3: [i ain't ne{ver been on the air ee oh pla:ne.]  
21 B1: {NO WONDER SHE WANTED TO GO TO NEW]=  
22 {A2 turns head toward A3 & smiles  
23 TV: )]} (oh my god) °  
24 B1: =YORK!]}  
25 }

### Transcript 2 – Competing for opportunities to comment

---

Then at line 15 a scene transition occurs, and the Real World roommate is shown at the airport saying goodbye to her aunt and her baby niece, to whom she says, “tell me *bah!*” or rather “bye” with a southern accent (line 17). Two of the participants then treat this as an opportunity to make additional jokes. In complete overlap with the TV dialogue, A3 makes a joke picking up the earlier backwardness theme, and mocks her southern accent (line 20). Just after A3 begins his joke, B1 comes in with another joke (lines 21 and 24) in overlap with A3 and the TV, suggesting perhaps that the girl’s family situation is driving her out of her small town to the big city. B1 not only produces her joke in overlap with A3’s but produces it with a loud voice (indicated by caps), to be heard over him. Thus we see the viewers doing a couple of things: prioritizing joking about the program over hearing parts of the program itself and competing for limited opportunities immediately following events in the program to comment on them.

In addition to talking around and over the TV program, viewers also used another practice for talking during a TV program in a group setting. In Phase 2, the viewers occasionally directed comments only to a subset of their fellow viewers, usually for only those in the same room, by producing them too quietly for those in the other

room to hear. Transcript 3 offers an example of such a *this-side comment* that develops into an extended conversation.

Transcript 3 (next page), also from The Real World, begins during a scene transition in which an image of the World Trade Center in New York briefly appears. After 2 seconds but still during the scene transition, A2 quietly comments that it is a “bizarre sight” post 9/11 (line 03). She produces the comments so quietly that only A1, who is sitting next to her on the couch, can hear it, but not loud enough for the viewers in room B to hear. After 4 seconds, A1 expands the topic by quietly commenting on whether such as scene might get edited out (line 05), and A2 quietly produces an agreement token (line 08). At this point, the program is still transitioning to the next scene with a montage of images from the city that are slowly narrowing in on the Real World apartment. The TV dialogue has not yet resumed. After waiting another 2.1 seconds, A2 expands the topic further by mentioning that an image of the World Trade Center was edited out of another TV show (lines 10-12). After another 2.4 seconds, A1 adds a different example of a movie preview in which the World Trade Center was removed. As he is describing the scene (lines 19-28), the TV dialogue finally begins (line 23). He completes his

description (lines 24-28), which includes a hand gesture depicting the Twin Towers, in overlap with the TV dialogue, and then they drop the topic to listen to the program.

This entire conversation is produced by A1 and A2 at a low volume and thus only for their room. A1's descriptive hand gesture further shows that he was designing his turns only for A2. Such *this-side comments* and conversations minimize disruption of the TV program by engaging only a subset of the

viewing audience. Because the verbal part of A1's description (lines 24 and 27) is spoken softly, it doesn't disrupt the ability of the room B viewers to hear the show. In addition to minimizing disruption through the loudness of their talk, A1 and A2 also do so by placing it in a scene transition. They appear to expand the topic cautiously by pausing for somewhat long periods after each expansion (lines 04, 09 and 14) in anticipation of the resumption of TV dialogue.

---

```

01      ((TV: 3-second image of World Trade Center, New York))
02      (2.0)
03  A2: °that's a bizarre sight°
04      (4.0)
05  A1: °i wonder if they would cut {something like that out°}
06                                     { A2 turns to look at A1 }
07      (0.7)
08  A2: °°yeahhh°°
09      (2.1)
10  A2: °.hhh they {took it out of the sopranos.}°
11                                     {A2 turns to look at A1 }
12  A2  °°didn't they. they took it from thee {uh::°°
13                                               {A2 turns back to TV
14      (2.4)
15  A1: °i remember there was uh: °
16      (1.2)
17  A1: °there was a preview for that spiderman movie°
18      (0.5) ((A1 & A2 look at each other))
19  A1: {that had
20  A1: {raises
21      (0.1)}
22  A1: hands}
23  TV: [{hello: / hi mo::m / how are you? /
24  A1: [{°the twin towers on it}{and they- they removed it}
25  A1: { traces Twin Towers }{ swirls hands around }
26  TV:  fi:ne, how are you: ]
27  A1: {and did some other cg stuff°}] =
28  A1: {R-hand chop & crosses hands }
29  TV: oka::y
30      (0.2)
31  TV: are you going to...

```

### Transcript 3 – Use of body orientation and whispers during side comments

---

Thus, in the preceding transcripts we see viewers using a variety of practices for coordinating their talk with the dialogue in TV programs to avoid or reduce disruption: fitting their talk in the gaps, using minimal responses, and producing quiet comments for a subset of their fellow viewers. However, we saw also that viewers are by no means constrained to avoid disruptions and will at times talk over the TV dialogue, as well as over

their fellow viewers, to compete for opportunities to make time-sensitive comments.

#### 3.2.2. Typology of comments

While we described *how* the viewers in our experiment coordinate their talk with TV programs and each other in the previous section, in this section we focus on content and offer a typology of the different kinds of comments viewers make.

Type of Comment	Examples edited from data
Content-based	"I love the way Homer dances in this scene."
Context-based	"Didn't Conan O'Brien write for the Simpsons?"
Logistical	"Could you turn up the volume?"
Non-Sequitur	"Did you hear about the tornado in Germany?"
Phatic	"Whoa!", laughter, gasps, groans, etc.

**Table 1 – Typology of comments observed during group television viewing sessions**

At a general level, we can characterize the comments exchanged by our participants using five broad types (Table 1):

- Content-based comments directly reference the content that is on or recently shown on the screen (as in the example in Transcript 1).
- Context-based comments are relevant to the show in its greater context, but perhaps not the specific episode or moment that is being viewed. Examples are references to the actors, past episodes, show trivia, etc.
- Non-sequitur comments are social exchanges such as asking about one's family, or talking about events unrelated to the TV program. These are usually more common in groups who already have some social connection with each other. They often take the form of side conversations (usually whispered, or at least toned down) between two participants and rarely follow the structure of the show.
- Logistical comments are relevant to the television watching experience, but are independent of the programming. Tasks like changing channels, adjusting volume, etc. must be verbally communicated to the group so that whoever has control of the set can respond.
- Phatic responses are almost involuntary reactions from the audience like laughter, gasps, groans, "Whoa!", etc. Nearly content-free (Schneider, 1998), phatic comments are not complicated interactions, but are vital to the social atmosphere in the room. Unlike other interactions which require turn-taking in order to make any sense, phatic responses have a "more the merrier" effect where additional "audience participation" adds to the sociable atmosphere. Phatic responses also do not require a pre-existing social connection to participate; both socially unfamiliar and socially gelled groups can share in a good laugh.

Our attempt at categorizing comments reveals that some exchanges are more disruptive to group television viewing than others. Non-sequitur, in particular, can negatively affect a group's experience since they often lead a small subset of participants to "talk over" the show about topics that may not be relevant to a majority of the group. In contrast, phatic responses do not affect the flow of the program and do not need to be so finely timed. They also scale well with the number of people: many participants can easily laugh (but not talk) at the same time.

To summarize, our observations reveal that *interactions between television viewers are tightly interwoven with the structure of the show they are watching*. In other words, the television program is both a resource and a constraint: it provides a rich conversational context that, to be properly exploited, requires the audience members to "read" the upcoming structure of the show in order to time their comments. Reaching a "flow experience" (Csikszentmihalyi, 1990) is important and viewers appeared reluctant to interrupt the show to communicate – the show itself has to be structured such that opportunities for communication exist. Moreover, only certain kinds of conversations contribute positively to sociability and private side-conversations can "derail" the audience from the show. The clearest contributors to sociability are phatic utterances and any kind of comment related to the show that elicits only a quick back-and-forth between a few of the audience members, such that the entire interaction sequence fits within the gaps provided by the television program.

### 3.2.3. Visual Interaction and Presence

During our observations, we also paid close attention to the body movements of our participants. One of the most surprising moments in our experiments was during a soccer game when a player almost scored a goal. Out of excitement, one member of the audience jumped out of his seat completely to react to the event, and crashed to the floor. Surprisingly, no one in the room shifted their focus from the TV to watch him. This illustrates another important feature of group television viewing, namely, that the interactions between the participants are visually peripheral: the televised content remains the visual focus. In most cases, it looks as if participants talk to the television itself instead of addressing any specific person by turning their head. The sense of audience presence is mostly conveyed via subtle cues in the room such as gestures perceived at one's visual periphery, movement in the room, and environmental audio (sound of people shifting in their seats, putting things on a coffee table, etc.).



Side conversations are a very different matter, however (Transcript 3). They are most often held with nearby audience members, and so as to establish a more personal connection, the participants will turn their heads towards each other to establish some visual contact (although as the conversation tapers off, the visual connection between the members deteriorates and the focus returns to the television). As we mentioned earlier, side conversations can negatively affect a group's experience but our observations of body movements show that they can be easily detected using the orientation of someone's face. This suggests potential avenues to dynamically move sub-groups in and out of the main audience – we discuss these possibilities and others in the next section of this paper.

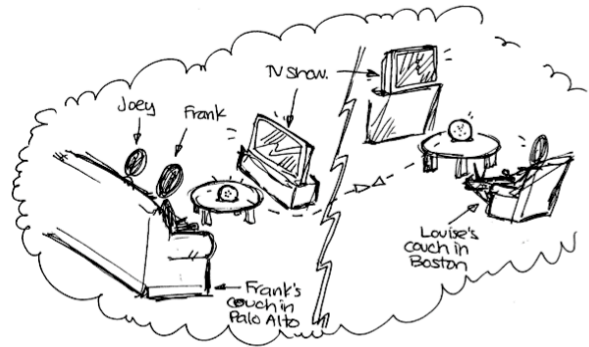
#### 4. Designing for Sociability

We believe that the subtle aspects of group television viewing we have just described can powerfully shape the design of future social television technology. Moving from observations to design, we will now briefly describe some possible features that could be implemented to support and encourage sociable, distributed television viewing. We want to emphasize that, while we plan to implement and test some of these features, our main focus so far has been on observing group behaviour, not prototyping. We hope, however, that the design possibilities we describe might inspire future research in this domain.

Our experiments using two rooms show that groups can socialize remotely while watching TV using a simple, always-on audio channel. Indeed, as we mentioned earlier, there were no major differences in behaviour whether or not our participants were co-located. Therefore, this basic communication infrastructure provided us with a foundation to start from.

Currently, we envision Social TV as a communication module that could be added to a PVR (e.g. TiVo), another piece of audio/video equipment (e.g. a receiver), or maybe the television set itself. It would allow viewers to establish connections with as many of their remote friends as they wish (probably using a mechanism similar to Chuah's (2002) "buddy surfing"), opening up a shared audio channel between these locations (using, for instance, voice-over-IP). Participants would communicate with each other simply by talking into a microphone that could be placed in the room (Figure 3) or, alternatively, on a small headset worn by each viewer. The main value of the technology, however, would be in the software available in each Social TV system. By processing

each participant's utterances and transmitting the appropriate mix of social audio content to all viewers, Social TV would act as a "clearing house" facilitating distributed television viewing.



**Figure 3 – A possible design for Social TV**

Based on the data obtained from our experiments, we believe such software would be particularly useful if designed to:

- Support the proper timing of social interaction during group television viewing;
- Minimize disruptions in the television program's flow;
- Isolate exchanges that are beneficial to the group from side conversations and non-sequitur;
- Allow viewers to move in and out of the audience smoothly;
- Avoid drawing viewers' attention away from the television screen.

Following these guidelines, we are considering the features listed below for Social TV system prototypes. It is important to note that, while we aim at simulating a conventional group television experience for distributed viewers, some of the features we describe will unavoidably change this experience. We have tried to minimize such disruptions and the amount of new skills required from the television viewers as much as possible. Note also that the list is far from exhaustive and we believe there is much room for future work:

- A preview of the oncoming show structure (for instance, a curve moving up and down depending on the amount of dialogue in upcoming scenes). Audience members would be able to call up this preview to "glance ahead" and accurately time when to start and wrap up their comments, so as not to interrupt the program. This would be particularly useful since, in the absence of the subtle cues given off by co-located viewers (e.g. heads turning), the show's structure becomes the

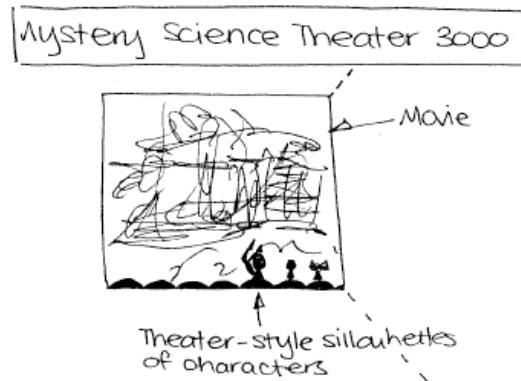
only reliable indicator of when to start and stop a conversation.

- Allow for variable-rate video when the Social TV system senses social interaction. When the audience keeps conversing close to the end of a break, our system could slow down and eventually pause the show if the conversation appears to continue after the show has resumed. This would prevent the ongoing television programming from stifling more developed social interaction. While this satisfies one design constraint (“encourage interactions”) it does, however, directly break another (“do not change the flow of the program”). User testing would allow us to explore whether or not the gain in sociability offsets the loss of continuity in the program.

Note also that such a feature offers the intriguing possibility of adjusting the length of advertising breaks dynamically. Since ad breaks are often used for conversations that have the potential to last longer than the break itself, our system could automatically keep adding more 30 seconds ads until the conversation is over. This opens up the door to alternative revenue streams and business models for content providers and broadcasters.

- Automatically isolate private side conversations. When audience members are not co-located, we cannot rely on body orientation to detect who is talking to whom. Technology developed in our laboratory, however, leverages insights from the same conversation analytic techniques we used earlier (Transcript 1) to automatically detect and isolate multiple, independent conversations in a shared audio space (Aoki et al., 2003). We plan to use such technology to facilitate interactions with Social TV.
- “What did they say?”: Often people want to hear every word of content but miss parts of it. Asking others about what was just said disrupts the flow of the program. Instead, our system will offer the possibility to turn on (temporarily and on demand) “delayed closed captioning,” so that others can read what was just said without interrupting the flow of the content.
- “Catch-me-up”: Since breaks are often used for bringing a new member of the audience up to speed, we plan to offer a “Catch Me up” button to facilitate the process. The Social TV system will automatically generate a one-minute visual synopsis of the show, displaying the scenes and moments that caused the biggest reaction from the participants (in a first approximation, we will use the general audio level in the room as an indicator of audience engagement with the show). New participants will be able to join the group,

quickly (and silently) watching the synopsis without disturbing other viewers. This would allow them to gain some shared background with the other viewers, which they can later use to participate in social interactions.



**Figure 4 – An unobtrusive interface to convey the presence of other viewers**

We have argued that Social TV should be as visually unobtrusive as possible, in order to keep the audience’s focus on the show. To convey a subtle sense of audience presence, we argue in favor of an interface building off of the visual scheme of the television show “Mystery Science Theater 3000.” It uses a movie theater metaphor to visually indicate the presence of other characters as they are watching movies. Similarly, the interface for the Social TV system could be a row of theater seats at the bottom of the television screen. When new members join in the Social TV session, their shadows would appear in the seats along the bottom of the screen. Whenever viewers want to judge how many other viewers there are, they can call up this visualization which remains briefly superimposed over the lower third of the television screen (Figure 4). In other words, the interface only appears when a “presence” event occurs, such as joining viewers, identifying speakers, or at the user’s explicit request. This is less distracting and uses far less screen space than full-bandwidth video showing each remote viewer, as in Huijnen et al. (2004). Of course, the theater interface does not convey as many social cues as full video, but Social TV compensates for this absence with the features described earlier.

Moreover, this interface could also be useful for selective sub-conversations. By selecting another user’s shadow, one could activate a more private, secondary audio channel where the two could exchange comments that the main group may find uninteresting or distracting.

## 5. Future Work

### 5.1. Technical improvements

Future work on the project will involve implementing the aforementioned design suggestions in a working system, and later evaluating their impacts and usability. To conduct further studies however, we first need to make sure that the quality and stability of the audio between non-collocated participants is enhanced. In our phase 2 studies, the quality of the audio transmitted between rooms left much to be desired, since it was plagued by echo and feedback issues that required significant time and effort to be minimized. Such time-consuming set-up was the cost of a quick and simple experimental configuration using off-the-shelf components but would be unacceptable in real homes. In fact the problem might be worse in the context of home deployments, since audio would be carried over the Internet and might suffer from greater latency than in our laboratory connection. Providing a pleasant social audio experience in the presence of loud and diverse program audio and latency is a challenge that requires attention to signal processing.

Beyond the transmission of audio between locations, we also plan to investigate the effects of providing a more specific directionality to the social audio so that it is easier for users to discern between the comments originating from remote viewers and the audio coming directly from their television. In our initial set-up both audio streams were sent to the same set of speakers and our users repeatedly pointed out that it could be quite confusing.

The issue of controls is also something we have yet to explore. We need to consider how our system would resolve control requests being issued from two (or more) distributed locations that receive a common video feed. One possibility is a virtual remote control that can only be used by a single “host” at a time. The host would control the flow of the show (e.g. pausing) while other controls (e.g. volume, brightness) would remain local. A marker on the host’s shadow (Figure 4) would indicate who is holding the remote, and other users could issue request to hand it over to them. This would further imitate the co-located viewing situation. But unlike co-located viewing, any participant should be able to wrestle the control back from any other viewer, in order to avoid some of the abusive behaviours described in previous research (Walker and Bellamy, 2001): “compulsive grazers” can be quite disruptive to a group’s experience, and the aforementioned research even describes users who claim to gain gratification from annoying others by controlling their viewing!

We hope that all of the above will eventually lead us to deploying our technology in more natural settings, as we discuss in our conclusion. We plan to provide families with Social TV systems so that we can see how daily television use is affected by the ability to share television viewing with remote friends and family members.

### 5.2. From synchronous to asynchronous use

In the introduction, we pointed out that television can foster two different kinds of sociability: direct (when viewers get together and watch the same show at the same time) and indirect (when viewers exchange comments after the show has already been seen by each, independently and at different times). While the former was the focus of this paper, the latter is rich in design possibilities and, in the course of our studies, some possibilities for asynchronous technologies emerged. In the interest of generating discussion in this space, we would like to briefly mention some possibilities offered by Social TV in the asynchronous realm.

It is possible to envision viewers running their Social TV device even when they are watching a show alone. The device would still record and track their reactions to the show but, instead of sending it to other viewers, it would store them for future use. As we mentioned earlier in our description of the “catch me up” feature, the reactions of the audience are rich in meaning and often indicate important moments in the show. Therefore, each individual viewer’s reaction could be used as a “tag” to mark important moments in a program. This could then be used by other viewers, at a different time, in the following ways (the list is, of course, not exhaustive):

- “My friends’ laugh track”: television shows often rely on edited laugh tracks to cue their audience about funny lines. However, these often feel highly artificial. Using the aforementioned tags, a solitary viewer could instead hear laughter only when his/her friends laughed when they watched the show earlier. The laugh track would sound much more natural and, moreover, it would be placed at the most appropriate moment – humor differs across viewers and it is probably better to rely on one’s friends’ reactions than the editing of the show’s producers.
- “My friends’ commentary”: commentary from movie directors and others is a popular feature on DVDs. Social TV suggests the additional possibility of recording a commentary about any show and sharing it with friends. A viewer would simply talk over the show and the voice stream would be recorded by the device. Each of his/her

friends, when viewing the same show at a later date, would have the option of turning on the commentary track he/she produced. In fact, this feature need not be limited to friends: viewers could offer their commentaries on the Web for anybody to download. This would emulate the popular “user reviews” offered on Web sites such as Amazon.com, but this time in the audio realm and with the added advantage of being synchronized to a video stream.

## 6. Conclusion

While counter-intuitive to many, watching television can be a very sociable activity. In this paper, we have explored how groups of television viewers interact with each other in front of the TV set. It became clear that television-mediated sociability is governed by a set of cultural practices and interaction rules, which apparently evolved such that joint viewers can simultaneously enjoy each other’s company *and* preserve the structure and pacing of the show they are watching together.

However, the pressures of daily life make joint television viewing increasingly difficult. To counter this trend, we suggested design avenues for a Social TV prototype inspired by our studies. We envision a system integrated into standard audio-video equipment (e.g. a TV set, a PVR) and allowing geographically-distributed viewers to communicate with each other using an open audio channel. Social TV would facilitate distributed, sociable television viewing by processing each participant’s comments and ensuring that they fit within the interaction rules we inferred from our empirical observations.

While we strongly believe in the value of this approach, it is important to note that our studies are not without limitations. First and foremost, our observations were conducted in a laboratory setting. While we tried as much as possible to emulate a real living room, the environment remains by nature somewhat artificial. Moreover, the participants were mostly recruited individually, without any attempt at reproducing the structure of social groups that most often watch TV together (e.g. groups of friends, family members). It could be that such prior relationships greatly affect interaction patterns in front of the television, but our data cannot shed light on this issue. More studies in “natural” environments (e.g. a participant’s own home) would certainly be enlightening.

The above could be most important when considering the form-factor of future Social TV systems and, in particular, the input devices it offers to its users. During our observations,

participants most often “talked to the television” when exchanging comments with each other, and changes in body orientation were rare and most often connoted private side-conversations. This, however, could be an artefact of our laboratory setting. Studies of domestic life (e.g. Hughes et al., 2000) have shown that home dwellers can be very mobile, conducting a variety of activities while moving about the house (e.g. alternating between television watching and trips to the kitchen to watch over a cooking dish). Television might also not be the exclusive focus of attention, with other tasks and devices asking for simultaneous attention (e.g. taking care of a child and watching television only peripherally). If these were the dominant patterns of use, Social TV would need to find a way to “follow its users around the house”, so to speak. This would imply decoupling the device from static A/V equipment and integrating it instead with another, more mobile platform (e.g. a headset, as mentioned earlier, maybe connected to a mobile phone equipped with the Social TV software). This way, users would be able to communicate about an ongoing show whether or not they are facing the television.

It is clear much work remains to be done to redesign television and make sure it continues to play a role in the ways people socialize and interact with each other. We hope this paper will help stimulate future research in this domain.

## Acknowledgements

The authors express their gratitude to the three anonymous reviewers who provided insightful comments and ideas that have been incorporated in the paper. We also thank our study participants for time and enthusiasm.

## References

- Alexander, A. (1990). Television and family interaction. In J. Bryant (Ed.), *Television and the American family* (pp. 211-225). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Aoki, P.M., Romaine, M., Szymanski, M.H., Thornton, J.D., Wilson, D. and Woodruff, A. (2003). The Mad Hatter's Cocktail Party: A Mobile Social Audio Space Supporting Multiple Simultaneous Conversations. Proc. ACM SIGCHI Conf. on Human Factors in Computing Systems (CHI '03), Ft. Lauderdale, FL, Apr. 2003, 425-432.
- Benford, S., Greenhalgh, C., Brown, C., Walker, G., Regan, T., Rea, P., Morphet, J., & Wyver, J. (1998). Experiments in inhabited TV. In *Proceedings of CHI98* (pp. 289-290). New York, NY: ACM.

- Casey, B., Casey, N., Calvert, B., French, L., & Lewis, J. (2002). *Television studies: The key concepts*. London: Routledge.
- Chuah, M. (2002). Reality instant messenger. In *Proceedings of the 2nd Workshop on Personalization in Future TV (TV02)*. Malaga, Spain.
- Coppens, T., Handekyn, K. and Vanparijs, F. (2005). Amigo TV, Alcatel White Paper, available at <http://tinyurl.com/9h68d>
- Crabb, P. B., & Goldstein, J. H. (1991). The social psychology of watching sports: From Ilium to living room. In J. Bryant & D. Zillmann (Eds.), *Responding to the screen: reception and reaction processes* (pp. 355-365). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Csikszentmihalyi, M. (1990). *Flow*. Harper Perennial.
- Glaser, B. G. (1998). *Doing grounded theory: issues and discussions*. Mill Valley, CA: Sociology Press.
- Hughes, J., O'Brien, J., Rodden, T., Rouncefield, M. and Viller, S. (2000). Patterns of home life: informing design for domestic environments, *Personal Technologies*, 4 (1) : 25-38.
- Huijnen, C., IJsselsteijn, W., Markopoulos, P., & de Ruyter, B. (2004). Social presence and group attraction: exploring the effects of awareness systems in the home. *Cogn Tech Work*, 6, 41-44.
- Lull, J. (1990). *Inside family viewing: Ethnographic research on television's audiences*. London: Routledge.
- Morrison, M. (2001). A look at mass and computer mediated technologies: understanding the roles of television and computers in the home. *Journal of Broadcasting and Electronic Media*, 45(1), 135-161.
- Oldenburg, R. (1989). *The great good place*. New York, NY: Marlowe & Company.
- Putnam, R. (2000). *Bowling alone: the collapse and revival of American community*. New York: Simon & Schuster.
- Sacks, H. (1992). *Lectures on Conversation*. Oxford: Basil Blackwell.
- Sacks, H., Schegloff, Emanuel A., and Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, pp. 696-735.
- Schneider, K. (1998). *Small Talk: Analysing Phatic Discourse*. Marburg: Hitzeroth.
- Schwartz, J. (2004). Leisure pursuits of today's young man. *The New York Times*, March 29, 2004. <http://tinyurl.com/csztu>
- Stone, G. P. (1981). Sport as community representation. In G. Luschen & G. Sage (Eds.), *Handbook of social science of sport* (pp. 214-245). Champagne, IL: Stipes.
- Subrahmanyam, K., Kraut, R., Greenfield, P., & Gross, E. (2002). New forms of electronic media. In D. G. Singer & J. L. Singer (Eds.), *Handbook of Children and the Media* (pp. 73-99). London: Sage.
- Sutton-Smith, Brian. (2001). *The ambiguity of play*. Harvard University Press.
- Walker, J. R., & Bellamy, R. V. (2001). Remote control devices and family viewing. In J. Bryant & J. A. Bryant (Eds.), *Television and the American family, second edition* (pp. 75-89). Mahwah, NJ: Lawrence Erlbaum Associates.
- White, E. S. (1986). Interpersonal bias in television and interactive media. In G. Gumpert & R. Cathcart (Eds.), *Inter/Media: Interpersonal Communication in a Media World* (pp. 110-120). Oxford: Oxford University Press.