***Title Page (including author names)**

VIDEO MEDIATED SOCIAL INTERACTION BETWEEN GROUPS: DESIGN GUIDELINES AND TECHNOLOGY CHALLENGES

Doug Williams
BT
Adastral Park
IPSWICH, IP5 3RE, UK
+44 (0)1473 647264
doug.williams@bt.com

Marian F Ursu
Department of Computing
Goldsmiths, University of London
LONDON, SE14 6NW, UK
+44(0)20 7919 7073
m.ursu@gold.ac.uk

Pablo Cesar
CWI: Centrum voor Wiskunde en Informatica
Kruislaan 413
Amsterdam 1098 SJ, The Netherlands
p.s.cesar@cwi.nl

Karl Bergström
The Interactive Institute
Box 1197
SE-164 26 Kista, SWEDEN
+46 (0)702 893544
karlb@tii.se

Ian Kegel
BT
Adastral Park
IPSWICH, IP5 3RE, UK
+44 (0)1473 642916
ian.c.kegel@bt.com

Joshan Meenowa
BT
Adastral Park
IPSWICH, IP5 3R, UK
+44 (0)1473 643365
joshan.meenowa@bt.com

VIDEO MEDIATED SOCIAL INTERACTION BETWEEN GROUPS: DESIGN GUIDELINES AND TECHNOLOGY CHALLENGES

ABSTRACT

This paper discusses results from research related to the use of television as a device that supports social interaction between close-knit groups in settings that include more than two locations, each location being potentially equipped with more than one camera. The paper introduces the notion of a *framing experience*, as a specific scenario or situation within which social communication takes place. It reports on the evaluation of some of the key attributes of social communication through semi structured interviews on the topic of a number of concrete framing experiences, with 16 families across 4 European countries participating. The issues identified through this study lead to the formulation of four design guidelines for the development of technology that could support such social interaction. The participants stress the importance of supporting excitement, engagement and entertainment, of high quality video communications, and of systems having the inherent flexibility of supporting and adapting to the unpredictable and reactive nature of human interaction and discourse. These findings were the basis for the statement of a number of hypothesized characteristics that a corresponding technology framework should have, and, in turn, allowed the formulation of a fundamental software architecture, whose main components are outlined in the paper. Finally the paper reflects on the impact the use of framing experiences, such as those described here, could have on strategy and policy for service providers and regulators.

1. INTRODUCTION

The television's affectionate place in the mind of society was, [Lull, 1980], [Kubey and Csikszentmihalyi, 1990] and probably still is, as a social family activity capable of binding significant fractions of a nation's audience to a shared experience. There is a strong feeling of togetherness among the members of a family gathered in the living room watching an engaging drama or the preferred soap. There is a strong feeling of connection with the whole nation when watching the national football team play live. There are many social moments created by the need to comment and share impressions, when face-to-face, about TV programmes. Television was and still is great for creating reasons for social interaction, but it was not itself a medium for social interaction.

Interactive television tried to address this shortcoming. The first generation of interactive television has imported a number of successful interactive web-based models, but has done this without observing that television watching is normally a shared activity whereas media consumption on a computer is usually an individual experience. Personalized (individual) television and obtrusive overlays over the television content are clear examples. However, over the last few years, we have seen an increased interest in research in interactive television as a shared experience.

For example, there has been work that considers group modelling for content recommendation [Aroyo et al., 2007], [Masthoff, 2004] and progress is being made on television experiences that can be shared in real time between households [Harboe et al., 2008], [Hemmeryckx-Deleersnijder and Thorne, 2008], [Coppens et al., 2005]. The work presented here builds upon this idea, but places a far greater emphasis on social communication and interaction, to the extent that the traditional, professionally crafted, TV content may be left out.

This paper reports on work that tries to define and to validate a range of new social interaction experiences between households that are mediated by the TV screen. It introduces the notion of *framing experience*, as a specific scenario or situation within which social communication and interaction takes place. Examples include parties, games and shared activities.

The remainder of the paper is structured as follows. It starts with a motivation and objectives section. It is followed by a section that outlines the method employed in the empirical research carried out on a number of concrete framing experiences. A summary of the findings of these experiments is then presented. The state of the art section, describing relevant research in social and interactive TV, communication in screen based games, video communications and composition, and interactive narratives and media sharing, creates the link with the technology section. This formulates a number of hypotheses about the characteristics that would have to be exhibited by technological frameworks that support the aimed social interaction and communication and then outlines the main components of a possible fundamental software architecture. The final section reflects on the implications that such proposals may have for government policy and industry strategy.

2. MOTIVATION and OBJECTIVES

Improving social communication is an inherently valuable goal, but this research is also driven by a simple profit motive. Even though the delivery of TV over the Internet is the focus of much corporate telecoms activity, communication, remains more profitable than media delivery [Odlyzko A., 2008]. The ultimate goal of this work is to explore new forms of rich social communication that will go far beyond the tools for social group communications such as Facebook[1] and Twitter[2] by employing the richness and transparency of video conferencing.

This work supposes that people are motivated to conduct social communication. This may seem self evident but it is also supported by motivation-theory, though, importantly, it is not intrinsically dependent upon the adoption or validity of any one model of human motivation. Debates on conscious and subconscious motives also matter little in this context [Forgas, 2005]; it does not matter whether actions are goal directed or subconscious. As long as there is a motivation within people to be sociable and to communicate, then the foundations of this work are supported.

We note that key models of human social behavior, such as the hierarchy of needs [Maslow, 1943], Alderfer's existence, relatedness and growth model (ERG) [Alderfer, 1969] and, in a more media-specific vein, the uses and gratification (U&G) model (Blumler and Katz, 1974) all identify as essential the need for social communication and interaction. Maslow speaks of love needs, Alderfer of relatedness and Blumler and Katz speak of

[1] Facebook: http://www.facebook.com

[2] Twitter: http://www.twitter.com

personal relationships, all of which will be nurtured by social communication.

In the models of Maslow and Alderfer we can argue further that, since the physiological (Maslow) or existence (Alderfer) needs are, in the developed world at least, met in abundance, social relatedness needs will be increasing. This is supported, at least in recent history, by the observation that there is significant consumer spend on meeting such needs. UK family spending analysis [UK Office National Statistics, 2007] reveals, for example, that the relative proportion of household expenditure on food (physiological/existence needs) between 1957 and 2006 has fallen from 33% to 17%. Meanwhile in the twenty two years to 2006 leisure spending (a category that includes a range of goods and services that exist to allow social activity) has increased from 12% to 19% of household income. Importantly for this work, this spending is largely confined to non-ICT based artefacts and services, wherein lies an opportunity.

The relationships needs cannot be met by an individual acting alone. They require interaction, at least with one other person and usually with a group of people. Love and esteem are received through communication, and the richest and most persuasive forms of communication, and therefore the form of communication we generally prefer when seeking to have such needs met, require a lot more than just verbal communication. The critical point is that in the communication of *feelings*, non-verbal communications are key [Mehrabian A, 1981].

ICT systems, particularly between domestic settings, are poor at meeting such needs as they do not allow people to see each other nor to hear clearly their intonation.

This research seeks to understand whether certain technological capabilities affect our ability to meet our higher order needs. Whilst technologies can be tested in isolation, their efficacy at enabling people to have their higher order needs met, can only be assessed within some kind of activity designed to allow social communication to take place. In this work we denote such activities as framing experiences and seek to evaluate them, as far as possible, within people's everyday lives. We seek to design and build a number of such framing experiences, and to evaluate their usage.

Our research intends to test a number of hypotheses related to the impact that certain technology developments can have on the way that people communicate and therefore (by proxy) to evaluate whether such technology developments can help people to satisfy their higher order needs.

3. METHOD

Television still plays a central social role in households[3]. It is posited that it should be possible to further exploit this role in the enhancement of social communication and connections between friends and family by developing new framing experiences that will help social interaction between groups. The methodology intended to test the hypothesis is user centred, with feedback and assessment from potential users being employed at many stages of the design and build process to challenge and refine the design. It is acknowledged that whilst a user centred approach is essential, users are less good at anticipating what they will like than recognising what they like once they have got it.

[3] Thinkbox www.thinkbox.tv "Growth of TV report". These UK based figures show the number of hours of TV viewed per day remaining effectively constant since 1993.

Our method is as follows:

- Suggest a range of framing activities that we believe may appeal to families as ways of enhancing social communication within and between households.
- Gain insight into the current social communication habits of families through a number of extensive interviews in four countries across Europe and to test reactions to the suggested range of framing experiences.
- Consider how technology developments might improve the framing experiences and to create a number of hypotheses that can be tested through the framing experiences.

3.1 Framing experiences

Five first draft 'framing experiences' were devised; each designed to offer people the opportunity to nurture relationships with friends and family.

The framing experiences attempted to offer a number of the following opportunities which were all believed to be intrinsic parts of social communication:

- For people to be able to talk with each other without the *explicit* use of a technological device and, even more, to have the sense of presence of the others even when not co-located.
- For people to be able to engage in group activities as if they were in the same physical space, whilst not co-located
- For people to share more about their lives with one another, when not co-located
- For people to be creative and to be able to share that creativity with others, when not co-located

The framing experiences include:

- A collaborative role playing game that allows teams from different households to enjoy a game in which the game play is dependent upon real time voice and video communication between households.
- A gentle familiar game designed to be played by older people that encouraged sharing of pictures and stories and provided ample opportunity for players to idly chat and enjoy each others company playing scant, if any, regard to the game.
- An application designed to allow young people in different households to show off their ability with tricks (like dance moves, football tricks or skateboard tricks) and to invite their friends to copy and learn from them and to practise together.
- An application that encouraged users to take part in incidental and indirect communication with the intention of helping friends to learn more about what is going on in each others' lives.
- An application designed to help people to develop personalised stories that they can tell to their friends and family, based on audiovisual material they have recorded themselves and on material recorded by others.

4. FAMILY INTERVIEW RESULTS

Early feedback from potential users is an essential part of a user-centred design methodology. In this section we report on the method of and results from a number of semi-structured interviews exploring the social communication habits of families.

4.1 Method

Sixteen families across four countries (UK, Sweden, Netherlands and Germany) were interviewed. We asked about their social communication habits, trying to gain insights into behaviour by

exploring with whom, how, when and where communication took place. We also explored attitudes towards communications technology. Towards the end of the interviews which typically lasted two hours, the families were introduced to the very generic descriptions of the pathfinder framing experiences being developed as described in section 3.1.

The households interviewed had children aged between about 6 and 25. The households tended to be of higher than average income and included a mixture of early adopters, the middle majority and laggards, according to Rogers's taxonomy on technology adoption. [Rogers, 1995] We explicitly sought households with long standing, deep bonds with another household.

Results are drawn from summary impressions written up from the interviews. The themes highlighted below are either those that were recurring in many households or those that tended to highlight and reinforce aspects of the social science review described earlier.

We report here our main findings with the express intention to inform our design of technology to help households nurture their relationship with their social contacts even when apart. We look at how the members of the households currently use games and playful activities and at how (whether) they let each other know how they are feeling. We discuss the reactions of our participants to our first draft framing experiences and look at the place audiovisual communications takes in the lives of our families, their use of multimedia and its social import. We draw on these conversations and feedback to infer some of the characteristics of our technology.

4.2 Results

The qualitative approach is not statistically representative, and includes divergent opinions. Nevertheless a number of common themes were raised in the interviews and from these a few hypotheses were drawn, relating to the potential impact of technology. These are discussed in the technology section of this paper. In spite of the sometimes frustrating nature of such qualitative investigations the insights from these interviews were a valuable source of information that helped guide the research. Furthermore, given the limiting costs of interviews, the qualitative study we report here best fits the stated goal of informing our designs.

4.2.1 Main themes

Following an analysis of the interviews a number of themes emerged as significant in at least three of the conversations.

- Play: playful activities and games emerged as a common part of the way families interacted in person. There was little evidence of enjoying games over mediated communication channels.
- Caring and communicating emotions; the families we interviewed talked about the way they gave care and communicated emotions using various channels of communication.
- Ease of use: many people were very keen to stress that new products would only achieve significant adoption if they were easy to use.
- Security and privacy: many people perceived as important the security and privacy aspects of

applications; these considerations influenced their behaviour and attitudes towards technology .

- Use of video: a number of families discussed using video communications but usually the experience was not found to be compelling.
- Media sharing; lots of our households described the way they shared media and discussed the role this had in their communication habits.
- Television and traditional shows; our households were still strongly tied to traditional media and liked to discuss TV shows with their friends.

4.2.2 Play

In general, the participants in our study referred to, and apparently valued, *play* as an activity that characterised and was enjoyed during social gatherings. However attitudes towards playing varied significantly between households, and interviewees in Sweden, in particular, expressed a marked lower interest in indoor games.

Delineating these playful activities in terms of degrees of flexibility or formality allows us to capture some of the characteristics of play which would best support our goal of nurturing long distance relationships between households. At the rigid end of the scale, we find the commercial video games aimed at consoles. For example, Wii games were very popular during meet ups and parties. Closer to the free-form end of the scale are made up quizzes/games which were sometimes reported as having been invented by a grandparent, sometimes described as involving *"running around a lot"*. In between these two categories lie puzzles and board games, whose rules can be negotiated and changed during play, and made up card games where only the physical format, the cards, is defined and the interaction can be completely invented though, in general, rules are not allowed to change during play.

All of these different sorts of activities were popular with our interviewees. The children would for the most sing the praises of the console games, and increasingly used terms such as *"silly"*, *"stupid"* and *"boring"* to describe the less formal games, and especially the free flowing, invented games. Interestingly, despite this, they seemed to have a much better recollection of the less formal games and appeared universally to have enjoyed playing them, with many a funny or memorable anecdote to tell.

We posit that although commercial console based gaming does provide for very enjoyable gaming, games focused on nurturing social interaction need to provide some flexibility in the game play to allow for non-task oriented intrapersonal interplay behaviour such as teasing, discussions, interruptions and rules and strategy evolution.

While our research has been concerned with the experience of the interaction, our participants were much more concerned about reliability of our proposed technology, even when some could envision such technology enabling games to be played across distance. Some household members also commented they did not see games as an end in themselves, but tended to play them rather as a convenient excuse for getting together physically; some of these family members were more sceptical of both the utility of playing and the ability to play family games on a network.

4.2.3 Caring and communicating emotions

Social bonding goes, of course, beyond play: as our participants noted, their interaction with their distant relatives is often quite

task oriented. In one case, a participant would be engaged in long phone calls and a long-running conversation via email to sort out with other siblings the arrangements for homes for an elderly parent. In another, one of our participants' brother would make a weekly Skype [4] call from a Latin American country to support her as she cared for their parents back in Germany. Several interviewees believed that technology might help not only in the coordination of such care but also in the provision of easy systems that might encourage greater communication or help keep an eye on those for whom they cared or for whom they felt a responsibility, typically a parent.

Broadcasting emotions, even to a select set of social contacts, however, did not appeal at all to the adults from our interview set. They found the idea of tweeting[5] messages with emoticons or expressing state of mind quite curious and alien, although several did comment that that was the nature of SMS messages or emails their mothers sent them. However, the concept was familiar to those of our participants who used social networking sites extensively. However, the fact that they could already do this on Bebo[6] or Facebook[7] meant that another technology that allowed the same functionality didn't appeal to them.

The idea of algorithms that judge their state of mind or identify their activity proved to be repulsive to the users, many reacting with an allusion to the "Big Brother" aspect of the technology.

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

[5] The act of post of posting a message on the Twitter social network.

[6] http://www.bebo.com

[7] http://www.facebook.com

4.2.4 Ease of use

Many families spontaneously cited basic tenets of good design as they considered necessary attributes of applications designed to encourage social interaction: simplicity, usability and reliability. Ease of set up, highlighted by this comment from an interviewee in Sweden, *"Ideally, you want it to work just like a radio; flick the switch and it is there, no startup, no nothing"* (translated from Swedish), was an especially strong theme. The 'hassle factor' was quoted by multiple families as a limiting factor on their use of their gaming consoles and webcam technology, and put them off purchasing such technology. Usability difficulties affected both younger and the older members, for example, they discouraged some younger family members (teenagers) from making full use of their mobiles. Reliability problems which included failure of appliances to connect to the home wireless network, low broadband speeds and lip synchronisation in video chat communications were often cited as a reason why a particular product or feature was not enjoyed or not used as much as might have been the case otherwise.

4.2.5 Security/privacy

Parents expressed concerns about security: about who has access to their information and especially their media - we encountered a participant who believed that his laptop had been taken over by a remote hacker and was thence cautious about his networked machines. Most parents interviewed believed they had at least some idea of what their kids were up to online, with awareness and thus possibly monitoring, decreasing with the age of their children. This ranged from "is vaguely aware of" to "controls" what their children are doing.

In some families, children shared and used their email accounts with their parents and one parent went as far as filtering and writing emails for her children aged 12-15.

The desire on the part of adults to exert some level of control over the digital lives of their children suggests strongly that applications targeting families should allow for "gatekeeper/ administrator" and normal user role. The gatekeeper would be able to define an application level policy covering who can communicate with whom, possibly when and for how long, in what way and share what sort of media.

4.2.6 Use of video

Only one of the families cited video based communication (in the form of Skype) as a regular part of their communication behaviour though many had tried it. A few families reported never having used video to communicate and saw very little utility in it. Some of our other participants voiced two main reasons for this: video gives correspondents a sense of each other's body language, and, it also allows communicants to get a feel of what is happening in the lives of the each other by providing background visual information. One family mentioned that video communication probably would be very suitable for group conversations, where it would be possible to see who was directing what to whom, who wanted to speak next, and so on; affordances lacking in a group audio-only conversation.

The emotional content enhancement, in line with Mehrabian's findings on non-verbal communication is typified in this quote from one of our participants as she justified her ownership and use of a webcam: *"When my brother [who lives in Latin America] was here last time he saw that I felt really bad and said that he wants to see me [using Skype], because you can tell a lot and this way you can see when somebody feels bad.*"

However, less keen users commented on getting bored by the webcam: we hypothesise this is at least partly attributable to the fixed view angle and the optical characteristics of the webcams used.

A second reported complaint in person-to-person communication via webcams relates to the physical presentation of the individuals: household members worried they might not come across in a visually appealing or socially appropriate manner via the webcam.

The use of audiovisual telephony for provision of background visual and aural information to create presence were reported in instances of absent family members who left their video chat sessions running *"all the time"* so that they could chat to one another and other. In another instance, one husband rings his family when away on business every morning to catch them during breakfast. His wife then puts the house phone on "hands free" mode on the kitchen table so her and her two children can chat to their Dad. However, the same participant, who professed to want to use video telephony *"with anyone, anytime",* also voiced his fear of his wife always knowing where he was, should widespread video telephony be available.

4.2.7 Household media sharing

Another important finding from our interviews was that many participants engaged in varied forms of media sharing as they felt that reliving memories and sharing experiences helped bring them (and other households) together. Teenagers reported showing *"random pictures of [them] messing about"*- images taken via their mobile phones - to their acquaintances via Bluetooth, and

"having a laugh" about these. Parents emailed pictures of the kids playing football to the grandparents, shared holiday pictures were communicated via Picasa or on disc or on Facebook, enabling friends and family to stay in touch with each others' lives.

All participating parents, if they shared media, would do so via communication methods they perceived as private: the so called private Picasa[8] album shares, email, via files on CDs or DVDs, and then only to trusted social contacts. There was a general reticence from the parents towards social networking sites. The children, however, tended to be far more open to the use of social networking websites, various instant messaging programs, and, in the less technologically adept households, tended to be the ones that led adoption of technology.

In all households we interviewed that had children, one parent would either be leading or controlling adoption of technology, or attempting to control what the children could do.

A number of parents reported photography as a hobby and would routinely edit their shared images. Their children, on the other hand, even if interested in photography, seemed less keen to manually edit the pictures, and declared a strong preference for automatic edits or relied on their parents. The participants would then discuss the incidents relating to the pictures later on with friends and family, on the phone or at the next reunion. Home videos tended to be watched far less frequently, with persistent remarks that they were "boring", although the young pre-teen participants appreciated them and were described by their parents as having *"worn the tape[s] down"* from constant viewing when much younger. Beyond concerns about privacy and data protection, the overwhelming UK participants we report on here reacted positively to the suggestion of automatic, intelligent edits of pooled videos of the same events and said they would happily use such a service if available via a web service.

4.2.8 Television and traditional shows

Our participants, particularly the adults, reported spending a lot of their time consuming media, especially television - be it traditional terrestrial broadcast as well as recorded PVR shows, or on demand media available via BBC iPlayer[9], 4oD[10], and YouTube clips and other audiovisual streaming services. Many of these shows, in particular soap operas and reality TV shows, would then be discussed at length with friends and family the next time they talked on the phone. Some shows were the *raison d'être* of phone conversations and one participant reported frequently watching shows while texting and phoning her friends as a running commentary on what was being viewed.

That experiencing media together facilitates the feeling of connection with others has been investigated in such work as ConnecTV [Boerjes, et al., 2007]. Our findings lend further support to the argument for further work in this area so household members can collate their experience of media with that of their social network and strengthen their bonds.

4.2.9 Cross cultural comparison

In the analysis some differences were noted between the answers from the families from different countries. Whilst it may seem tempting to frame these differences as a cross cultural comparison, because of the statistically unrepresentative nature of the study, any attempts to do so would be inappropriate.

[8] http://picasaweb.google.com

[9] http://www.bbc.co.uk/iplayer/

[10] http://www.channel4.com/4od/index.html

5. STATE OF THE ART

This section introduces the state of the art in a number of relevant technology areas:

- social and interactive TV – this is discussed because the framing experiences described here use the television (as an artefact) in the development of effective social interaction
- communication in screen based games – this is discussed as the framing experiences often build on the idea that playful activities are central to social communication and we seek to understand how communication has so far been handled in screen based games
- video communication and composition – this is discussed as a premise of the work is that the ability to see people is central to effective social communications. The section looks at experiences and usages of video communications and of the way video communications session and other media may be composited on the screen
- media sharing – this is discussed as we observe that sharing showing and discussing media is a common framing activity for social interaction and the framing activities developed will build on this behaviour
- interactive narratives – this is discussed as we postulate that users may enjoy the ability to use their own media to tell personalised stories to their friends; this is being tested by building pathfinder framing experiences.

5.1 Social and interactive TV

The research field of interactive digital television is being transformed into a study of human-centred television [Cesar et al, 2008a], in which television viewers become active users with communication and (re-)distribution capabilities. Unlike the first generation of digital television systems, which mostly focused on the concerns of content producers and device manufacturers, currently there is a wave of research that tries to leverage the role of the user in the distribution chain. Such development represents a step towards the innovative framing experiences foreseen in this article, which exploit social communication opportunities such as presence-awareness and communication capabilities.

As an active user, the television viewer might want to communicate with others while watching [Chorianopoulos and Lekakos, 2008], [Ducheneaut et al., 2008], to leave notes and comments for friends at specific moments of a television show [Nathan et al., 2008], and to share enriched fragments of multimedia content with others [Cesar et al, 2008b]. For example, TV-based services like Alcatel-Lucent's Amigo TV [Coppens et al., 2005] allow users to watch broadcast TV together (when apart) and to augment their watching experience with voice chat, messaging and the use of emoticons. The Social Television project [Metcalf et al., 2008] by Motorola, apart from synchronous communication mechanisms, provides an unobtrusive awareness system based on ambient devices. The final goal of these approaches is to provide enriched communication between separate parties, when watching television content. Similar developments are occurring in communication-oriented systems, where systems such as Zync

[Shamma et al., 2008] extend traditional instant messaging capabilities with a synchronized watching experience.

The framing experiences proposed in this article extend current work on social interactive television by incorporating the communication capabilities within the shared media experience across different households. While previous research considered a direct communication link (e.g., text chat or audio chat) as a meta-activity, rarely related to the content being watched, in our research media content and communication are orchestrated and composed in a coherent manner; as a single unit. At the same time, first generation social interactive television systems normally did not consider collocated experiences. Even though people were connected across distances, the basic assumption implied that only one person was in front of the television set at a given time. Our work, on the other hand, pays special attention to collocated multi-user settings, connected to other collocated multi-user settings. Hence, innovative work on audiovisual cue detection (e.g., detecting the person who is talking at a given time) is an intrinsic part of our suggested framing experiences.

5.2 Communication in screen based games

The communication between players in modern multi-player, distributed computer games comes in several layers depending on how far the gaming activity has progressed. This communication varies in sophistication with different games and gamer cliques, ranging from asynchronous messages (email, forum posts), simple near real-time text messages (chat, instant messaging) to full real time, duplex audio communication. It is important to note that not all the layers are present with all games and all gamers, naturally.

The first layer concerns communication *surrounding* the mediated event (i.e. the playing of the game itself) where gamers talk about the game, discuss strategies and memorable moments. This kind of communication is typically in the form of blog entries and comments, forum posts and email.

The second layer of communication takes place *right before* the event itself, as players negotiate the particulars of the game. This usually takes place in the "lobby" of the game, via real-time chat. Players-to-be discuss parameters of an upcoming game, such as difficulty level, specific map, rules and participants.

During the gameplay, methods of communication become more varied, and for many games and gamers this means real-time, duplex audio communication. Support for this is rarely provided by the game software itself and the gamers instead utilize third party solutions, such as *Ventrilo*[11], *Teamspeak*[12] or *X-box Live*[13]. In this phase, the content of the communication is mainly focused on the gameplay itself; giving orders and heads up and discussing strategy, for example. Real-time text messaging is also often used for the same purpose, but is generally considered inferior, since it is often difficult to game and type at the same time.

Games that progress over several "rounds" also see significant *between-round* communication, as gamers await the next round. Part of the success of acclaimed *Counterstrike*[14] is said to stem from the fact that gamers do not respawn immediately after dying, giving them time to discuss the game in between, contributing to the culture of the game.

After playing a game, gamers might linger in the lobby of the game to discuss the event that just took place. Analyzing what

[11] Ventrilo Client and Server (1999-2007) Flagship Industries Inc
[12] Teamspeak (2008) TeamSpeak Systems GmbH
[13] Xbox Live (2008) Microsoft Corporation
[14] Counterstrike (2004) Valve

went wrong or right, picking out specifically memorable events, commenting on game balance and more.

The more of these layers of communication that a system can facilitate, the greater chance users will continue to play for a greater period of time. *Wii Speak*[15] seems to strive in this exact direction as it facilitates many different layers of communication before, during and after a game.

It is clear that if gamers perceive a need to communicate and have the means for it, they will do it, even though the game software or game rules might not support it (or even try to forbid it). Players can always bypass the provided software and rule set if they feel it is necessary. Thus, any system designed for social gameplay using TVs must take this into account.

5.3 Video communication and composition

Videoconferencing is now over 40 years old and has found a degree of commercial success with dedicated hardware solutions primarily for business applications. These range from traditional standard definition systems to high end 'mirrored' telepresence environments such as those offered by *Halo*[16] and *Telepresence*[17], the latter almost always requiring a dedicated room in which the environment (for example lighting) can be carefully controlled. High-end systems also guarantee quality by using dedicated high-bandwidth networks and expensive components, and are offered as a managed service. Recently, manufacturers have begun to announce[18] home telepresence systems which will operate over the Internet and use some existing home components, such as a high-definition TV. While these systems will initially carry a high price ($10,000 according to some estimates), they give a clear indication that high-definition home videoconferencing could become viable for the mass market within the next 5 years.

Several challenges face developers of high-quality videoconferencing systems which will connect via the Internet. End-to-end delay, packet loss and jitter will all vary according to network conditions, and their effects can be exacerbated when a multi-point architecture is required to accommodate more than two locations communicating simultaneously. Han [Han et al., 2008] provides a comparison of delay and bandwidth requirements for different multipoint topologies in the context of domestic videoconferencing. Arguably the most promising recent development is the implementation of the new H.264 Scalable Video Coding standard [Davis, 2006] within Internet videoconferencing applications. The ability to dynamically control quality by the use of enhancement layers can accommodate changing network conditions, and delay is significantly reduced by removing the need for decoding in a multipoint control unit.

At the other end of the scale, video chat has seen a significant increase in popularity, fuelled by free applications such as Skype and the ubiquity of webcams and laptops with integrated cameras. At the end of 2007, 23 million people were using video chat services in the USA alone. Standards such as SIP [Rosenberg et al. 2002] and IMS[19] have opened up new opportunities for fixed-mobile convergence but the majority of the new applications they enable are focused on person-to-person communication only. At the same time, there is growing evidence [Harmon, 2008] that

[15] WiiSpeak (2008) Nintendo

[16] Hewlett-Packard Halo Collaboration Studio

[17] Cisco TelePresence conferencing suites

[18] http://www.crn.com/networking/208801088

video chat is being used by geographically-distributed families to meet higher order needs, but the deficiencies and one-to-one nature of the technology are also evident.

The 'digital home' is now a commonplace term, and organisations such as the Digital Living Network Alliance (DLNA)[20] are building interoperability standards for consumer equipment which seek to harmonise media servers, players and controllers within the home environment. However, the current focus of their efforts is on the delivery and rendering of fixed, linear media assets such as Video-on-Demand and music. The UPnP AV standard[21], a key building block of DLNA, specifically excludes two-way interactive communication such as videoconferencing.

Significant work has also been carried out on *integrating composition formats*, or languages for specifying the temporal synchronisation and spatial layout relationships between several digital media elements. Both the W3C's SMIL[22] and ISO MPEG-4 are examples of such languages. However, neither is yet capable of describing the multi-layered compositions which are required when a combination of live audiovisual streams and pre-recorded content is to be simultaneously orchestrated for (potentially) multiple display devices in multiple locations.

[19] SIPK, 3GPP IMS Specification List, available at *http://www.sipknowledge.com/IMS_Specs.htm*, 2008.

[20] DLNA, Overview and Vision White Paper 2007, available from *http://www.dlna.org*, 2007.

[21] UPnP, AV Architecture:1 (Approved Design Document), 25 June 2002, available from *http://www.upnp.org/specs/av/*, 2002

[22] SMIL, Activity Statement for W3C Synchronised Multimedia Working Group, available from http://www.w3.org/AudioVideo/Activity.html, 2008.

5.4 Media sharing

Many high-profile and popular websites have been established to host user-generated online video and to facilitate social interaction among communities of creators and consumers, including *YouTube*, *MySpace*, and *Ovi*. Their new models for publishing and distribution are characterised by the two-way communication between creator and consumer which enables content to be discovered, annotated and even 'remixed' by a third party – hence the term 'conversational media' [Battelle 2006].

The majority of conversational media content can be classified as 'self-expression' between an individual and a potentially wide audience. The popular websites are little-used by families and small social groups for the collection and sharing of media which is personal to them. While they do provide seamless upload, transcoding and delivery for content, their simple free-text descriptions do not guarantee a consistent approach to its annotation, and hence their utility as a creative resource. This means that considerable effort is usually spent authoring complete stories prior to upload using consumer-focused editing tools (such as *Apple iLife*, or *Microsoft Windows Movie Maker*), or to 're-mix' stories from multiple uploaded fragments.

A subset of user-generated online video sites offer more advanced editing features as a replacement for desktop software. However, neither these nor their traditional equivalents provide help in the creation of a good narrative or the ability to automatically tailor that narrative to different viewers.

5.5 Interactive TV narratives

Interactive TV narrativity is a subset of interactive television in which the (active) viewer can influence the programmes (the stories) that he receives. The relationships between this new form

of television and the framing experiences are evident. First, social communication is normally framed as a narrative. Interactive narratives allow end users to "converse" with the narration (or with a virtual storyteller). Our framing experiences require this kind of communication, but take it further: interactive TV narratives, or interactive TV programmes, in their current understanding, are assumed to be pre-authored by experts and subsequently delivered in an interactive manner to the active viewers; the framing experiences aimed in our research do indeed require interactive narratives, but they cannot be pre-authored by experts: they have to be compiled automatically, in real time. Second, recounting aspects of a social (or fictional, for that matter) system with moving image is a (long) tradition of television. Social communication as envisaged in this paper, mediated by moving image, will have to build on these grammars, adapt and use them to ensure the naturalness of the communication.

Interactive TV narratives have been developed in established TV genres, such as drama, documentary and news, and notable examples include: *Façade* [Mateas and Stern 2005] and *Accidental Lovers* [Tuomola et al., 2006/2007]; *Terminal Time* [Mateas et al. 2000], *Vox Populi* [Bocconi et al. 2005] and *A Golden Age* [Zsombori et al. 2008]; and *My News And Sports My Way* [Larsson et al. 2008]. They are attempting to move into a space of agency, whilst preserving the quality of the storytelling, a move which is complemented in the arena of agency-based screen media (games), which are gradually incorporating elements of narrativity [Ursu et al. 2008b].

Agency centred screen media (games) has started to incorporate narrativity. Screen media narrativity, embodied by television programmes, has started to move towards viewer agency. Finally, in some forms of game-play we note that a created world (the game) is present along with the ability to speak to and see each other. The framing experiences described here could be regarded as the centre of convergence for these developments.

From a different perspective, we are exercising our privilege to creativity when we tell stories to each other, even when we talk about trivial aspects of our lives. When the communication happens face to face, our creativity is instantaneous; it is more or less an improvisation act. However, if the communication is to be time-shifted and carried by screen media, our creativity in telling such stories has to be more explicit. The model of interactive screen media, best represented by ShapeShifting Screen Media [Ursu et al. 2008a], represents a good paradigm for this

6. TECHNOLOGY

The results of the family interviews described above support four hypotheses which guide thinking about how current technology can be extended to meet future users' needs for social, communication-based TV experiences.

6.1 Hypotheses

The key themes emerging from the family interview results together with an understanding of the state of the art lead us to the following hypotheses about the obvious characteristics that would be exhibited through the technology framework used to deliver the framing experiences:

- Support for engagement, excitement and entertainment: To support users' willingness to take part in playful activities, the system must be designed to support a strong sense of engagement between the parties, to enable the generation of excitement and deliver

entertainment. Systems with such capabilities could also go some way to overcoming a common criticism of current video based communication – that the experience was boring.

- Flexibility: Systems need to be able to support the dynamic and almost chaotic nature of real life playful interactions and to be able to support less rigid forms of storytelling that is experienced between friends and often supported through media sharing (e.g. photographs).
- High quality video communications: To overcome some criticisms of poor shaky video experienced in current video communications systems, we should deliver high quality (TV like) images recognizing that such images will aid in the delivery of care to loved ones and in reading and communicating emotions
- Control and usability: A number of issues such as privacy, security can only be delivered if the system has control and usability, and these characteristics are unlikely to emerge except through a user-centred design approach.

These hypotheses are discussed in more detail below.

The mapping between the key themes and the hypotheses is shown in Figure 1

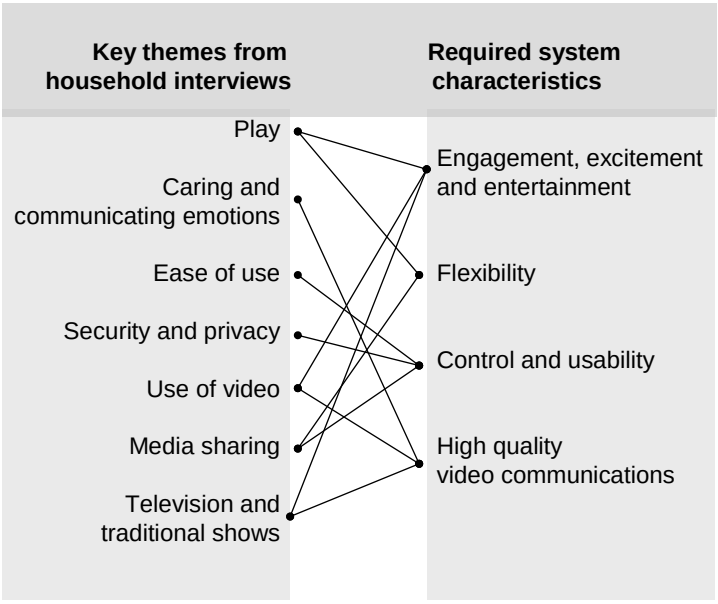

Figure 1. The relationships between the required characteristics and the key themes from the family interviews.

6.1.1 Engagement, Excitement and Entertainment

While several families appreciated the value of good non-verbal communication in providing, for example; background information, visibility of body language and opportunities for group to group communication, some of those who had experience with webcams cited the experience as boring. We interpret this attitude as a vote of no confidence in the *current* technology.

In the traditional representation of professionally authored moving image, such as that on television, we are the fortunate viewers of a craft skill in video storytelling that has mastered continuity editing and that allows us to naturally engage with the pace and excitement of the story being told. This quality motivates end users to spend significant time on consuming media.

When it comes to video communication, the aforementioned skills in representation are absent. The point of view is either static or slowly wanders to capture the current speaker by controlling a PTZ (pan-tilt-zoom) camera to focus on the source of the voice, perhaps aided by detection of the speaker's face. Instead we propose that audiovisual communication is part of a wider framing experience than a simple peer to peer fixed video 'window'. There may be an arbitrary number of people in each video input and an arbitrary number of different groups of people sharing the framing experience, and furthermore the on-screen representation of an application, such as a game, may also be shared by the groups. Communication between these groups of people should be natural and require no *active* intervention from them. It should also provide a quality similar to that of professionally-crafted TV programmes, which have been proven to be able to engage the viewers in their space as if the space were real. This would ensure, we believe, much better levels of engagement in communication between the groups than the current blunt and low quality videoconferencing systems, which, in turn, will enhance the excitement and entertainment generated by the communication.

6.1.2 Flexibility

As described above, our family interviews discussed playful social activities between family members and introduced several framing experience concepts. The families' reactions suggest that framing experiences should allow flexibility for interpersonal interactions which are beyond the necessary communication required for game play. While this flexibility can easily be accommodated in co-located casual games, it is much more difficult to manage for any activity which is distributed between multiple locations. It should be an integral part of the craft skill of storytelling which we propose and thus it should be embodied in the audiovisual communication.

Another requirement is to allow for different gaming activities to be included in the overall system, to accommodate the different habits and tastes in the social activities which bring close groups of people together.

6.1.3 Control and Usability

Many families placed a high priority on the properties of control and usability, which are often in conflict in today's communication applications. Control was important within families, for example to ensure children were viewing appropriate content. It was also important beyond the home when people expressed concerns about the privacy of audiovisual communications and sharing information about their activities and emotional state. This was further extended to the sharing of media, which was generally done through methods perceived as private. However, usability and the 'hassle factor' were also key influencers on the adoption of new technologies. While people appreciated the need for control, they also demanded a system which can be invoked with a single switch; these are contradictory requirements only in appearance and could be met together by a well designed and implemented system.

Our hypothesis for control and usability is not based on a new technology solution, but on principles of good user-centred design:

- Consistent responses to user interactions: We are proposing a system which is inherently complex. It is essential that users' interactions with the system are managed consistently, especially when some will relate specifically to

communications and some specifically to an activity such as a game.

- Simple metaphors: The process by which a social, communication-based TV experience is set up must be managed logically and without visibility of technology. Users must be able to discover each other's availability and enter a shared activity using uncomplicated metaphors.
- Simple tools: We are also proposing to help people create their own personalised stories from a shared collection of media. The state of the art in personalised interactive narrative technology [Ursu et al., 2008a] is focused on creative professionals. These existing tools will be simplified to the extent that they can be used by groups of families and friends.
- Appropriate devices: The most appropriate devices and applications must be chosen to manage information and interactions with the system. For example, a mobile device may be used to store private information and content, and existing social networks may be used to reveal personal feelings and interests.
- Accessibility: The system must be able to adapt its user interfaces to the accessibility needs of a wide variety of individuals at the same time.

6.1.4 High-quality, reliable audiovisual communications

Several people interviewed highlighted the importance of reliable audiovisual communication, and when using webcams had observed problems common to all video chat sessions: errors in lip synchronisation, low frame rates, poor video quality and occasional loss of connection, as well as concerns about security of the communication channel.

We propose that the state of the art in high-quality videoconferencing should be brought into the domestic environment and made to operate using high-definition televisions and over contended broadband networks. Further, this high quality experience must be maintained even when additional content, and the representation of applications, is combined with live audio and video streams.

6.2 Technology Components

Systems built to successfully deliver the framing experiences in section 3 should exhibit the characteristics listed in section 6.1 above. In order to do so we infer that user-centred design principles must be employed to ensure that users experience the control and usability they desire. In addition a number of technology problems need to be solved. These are listed below:

- automatic orchestration of the audiovisual communication
- multimedia interpretation
- multimedia composition and delivery

Each of these, and orchestration in particular, comprise major research challenges. The following sections describe in more detail how these technology components are being developed.

6.2.1 Orchestration

Interaction or communication orchestration refers to the automatic reasoning processes that control:

- what is captured by the cameras in each location

- what is edited for presentation on the TV screen and for reproduction through a (spatial) audio system at each location..

In the compilation of a live TV transmission, the director has a number of views of the live event, through different cameras and microphones, but he also has access to an archive of previously recorded material. He decides how to combine information from all these sources in the broadcast programme in order to best recount the main event. Interaction orchestration has a similar function, but in the context of the communication between the friends and family interacting in different framing experiences. It ensures that each participant in the interaction has the best perception or view of the others, such that the interaction seems as natural as possible. Orchestration could be regarded as *automatic* or *virtual directing*. Continuing the metaphor, orchestration refers to all the decisions that cameramen, director and possibly editors take when composing a programme recounting a live event – namely, which parts of the "action" to capture (audio and video), which to select for inclusion in the main programme and how to edit them together.

Orchestration has a more complex task compared to that of a programme director, in that it does not compile a single narrative thread, as in a broadcast programme, but has to compile a separate thread for each of the participating sites in the interaction. The viewers of a live TV programme cannot (currently) influence in real time what they see. In the case of social interaction, the *viewers* are the same as the *actors* and thus they influence, both directly and indirectly, each other's behaviour, and therefore what they see and when they see it.

Orchestration is therefore, in technical terms, about

1. eliciting logics of interaction facilitated by audio and video, and
2. representing them in computational formats which could be interpreted automatically by software.

The first task is not straightforward, as there is no dedicated body of knowledge upon which we could draw. Programme makers know how to compile representations for live events, interviews, documentaries, etc., but not for small groups of people interacting from different locations for pleasure and fun. Anthropologists and psychologists look at the forms of communication between people, but not from the point of view of how to record and convey the most important aspects of the audiovisual communication. Game designers look at how to enforce game rules, but they do not focus on the side interaction, i.e. that which is outside the game logic. Even when the objects of the game are the people themselves, the game logic remains in focus. The main types of intelligence that need to be elicited for orchestration are:

- what audiovisual information to capture – where to point the cameras and on what to focus; the captured content is used for live communications (in real time) and well as being stored for later processing and use.
- what audiovisual content to select from each site – which fragments from the continuous streams generated by cameras and microphones to extract and, possibly, which parts of the selected video and audio to further extract (e.g. only a part of the screen from a HD moving image); selections can also be made from previously recorded and stored material.
- how to edit the extracted parts for each site participating in the live interaction (depending also on the available delivery

devices, such as number and sizes of screens and number and positions of speakers)

- how to edit for later use (i.e. not for real time communication/interaction).

We aim to create a declarative representation language for the expression of *interaction scenarios*. These are statements that describe what the system should do in specific situations of particular framing experiences – i.e. what the cameras should focus on and what and how to edit in each location.

Orchestration relies on and is constrained by both *multimedia interpretation* and *multimedia composition*.

The aimed communication/interaction could be real time or could allow for delays and iterations. Through orchestration we aim to provide for all these possibilities. Constructing such solutions to ensure best communication between sites is the subject of this research.

6.2.2 Multimedia Interpretation

Multimedia Interpretation refers to the extraction of textual information possibly accompanied by audio and visual objects, generically called *features*, from captured media objects (and possibly ambient devices).. Two distinct challenges have been identified,

1. the extraction of audio and visual semantic cues to enhance and capture the experience, and
2. the detection of events and trends using the audio and visual cues.

In the first challenge, we expect robust modules to be capable of extracting speaker identity, speaking activity and to identify the types of interaction taking place (e.g. monologue, discussion) as well as other body activity (e.g. head gestures or focus of attention). [Ricci, 2009] In the latter challenge, events are defined on a short-term temporal scale and include the detection of turn-taking patterns and the recognition of conversational events (e.g. direct exchanges, general discussions). Trends defined at a medium and long temporal scale include the estimation of participant engagement in the activity or the recognition of the group interest-level (the degree of engagement that the members of the group collectively display during their interaction).

The outputs could include "the person now speaking is … and is captured in the rectangle positioned at … and of size …", "there is an intense or excited conversation between … and … (which could further be interpreted into an argument)" and feed the reasoning processes subsumed by orchestration. Only what can be captured through media interpretation can be further reasoned about and orchestrated.

6.2.3 Multimedia composition and delivery

The challenge of interaction orchestration between two or more distributed sites, whether in a synchronous or an asynchronous mode, creates a corresponding challenge in realising the media experience continuously being computed and described for each participant. Today, the state of the art has furnished us with a toolkit of components, many backed by international standards, which individually provide certain aspects of the solution.

Three types of media composition are identified:

Base Composition: This is the composition of a number of autonomous media assets into a structured presentation, such as one of the framing experiences. This content may need to be reused many times by various collections of participants. A base

composition may be provided by an external application developer, or it may consist of 'pure' initial media objects, such as home videos or digital images, which are defined by one or more people and shared by many external users.

Interactive Composition: There will often be the need to perform dynamic compositions of sets of base media objects, possibly augmented with additional media that represent the encoding of a user's action within the framing experience. Compositions that result from a particular viewing session must either be dynamically captured and encoded, or come from a library of possible actions.

Third-Party Compositions: Whenever a particular action is taken in the framing activity, it may be necessary to record one of more observations or reactions from other participants. Such commentary is a collection of ordered media objects that are related to the other composition types described above.

We do not explicitly seek to develop new content rendering software or hardware, or media encoding formats. Instead we will develop an assembly of state-of-the-art components described above which is capable of dynamically combining live streamed content (for example, from cameras) with pre-compiled compositions. Furthermore, the composition and rendering component will need to be scalable, so that it can support a large number of simultaneous sessions, each of which may contain multiple media objects. Such complexity may favour a hardware-assisted solution.

It is recognised that the perceived quality of framing experiences will be significantly affected by the quality of the audio reproduction for each user. This is relevant to an asynchronous experience, in which a user receives a composition made from pre-existing media components, because the combination of music, effects and speech has a critical creative role in defining the mood and continuity of a narrative. However, it is even more relevant to a synchronised experience constructed between multiple physical environments because social interaction between two or more individuals relies heavily upon them being able to hear each other's speech clearly and consistently. The goal of this research is to enable this communication while the capture and reproduction hardware remains as unobtrusive as possible. It is generally accepted that an effective audio subsystem can improve the perceived quality of a rich media experience in which the visual elements are subject to delays or reduced quality.

We seek to advance the state of the art throughout the end-to-end audio subsystem by integrating tools for echo control, improved error concealment and spatial audio into an MPEG-4 Enhanced Low Delay AAC (ELD) encoder and decoder. The Spatial Audio Object Coding (SAOC) technology to be used adopts a parametric approach to describe audio objects which dramatically reduces transmission bandwidth and enables flexible reproduction. Furthermore, special algorithms for bit stream mixing inside the codec domain will be developed which reduce complexity and delay, and increase quality of multi point connections. New frame loss concealment algorithms, which make use of codec signal representations, will also be deployed to minimise delay and preserve audio quality.

As can be seen from this discussion, the challenge of multimedia composition and delivery can be broken down into a number of unique problems which must all be addressed in order to realise the end-to-end delivery of complex, interactive audiovisual services between homes.

7. USAGE AND POLICY IMPLICATIONS

The framing experiences described in the paper describe applications that exemplify "convergence". In this case they represent possible modes of convergence between television and telephony. They suggest rich communication between people that know each other well, based on the combination of different technology artefacts (screens, routers, microphones, cameras) through an application that incorporates both media and communications elements. Such usages have significant implications for industry strategy, and for policy makers.

One of the key elements of the usages is that existing pieces of consumer equipment are used in combinations that enable new experiences (services) to be supported e.g. using the TV and a video camera as a video conferencing device. The combination of disparate technology devices usually requires standardisation and industry has been successful in defining standards which have allowed it to grow the market very successfully. Key examples include the MPEG standards for picture encoding and transmissions standards for television. More recently the industry has come together to create the Digital Living Network Alliance (DLNA) which according to the web site[23] (as of September 2009) is "a global collaboration of 245 most trusted brands working together to create the home entertainment you've always imagined". The DLNA recognises that users will expect different devices (TV's, media servers, PDAs, mobile phones, PVRs etc.) to work together seamlessly. Apart from the obvious challenge of achieving a workable consensus between the large number of brands another key challenge for the DLNA is to correctly anticipate the entertainment "you've always imagined". Consumers are not good at imagining new services, but they are very adept at harnessing new technology in ways valuable to them that the industry had not necessarily imagined; telephony and SMS messaging are relevant cases in point. Currently the DLNA imagines usages that are *not* real time, and that predominantly envisages the behaviour within one house and the movement of and access to media within that one house. This is changing; emerging work in the DLNA such as "remote access" and a working group looking at the requirements of content service providers both consider use cases that include the delivery of media from outside the home and access to your media from another home.

There is, as yet, no substantive effort to support the rich media real time communications activities such as have been described above. With no such effort only proprietary solutions are likely to achieve any level of success and it seems unlikely that these will achieve the levels of mass penetration required for such services to really become commonplace.

Brands, and collections of brands like the DLNA, have an opportunity to define strategies that will accommodate the real time framing experiences described previously. The difficulty is, as ever, imagining the future. Are such usages really attractive? If so, how can the industry, and in particular service providers, profit from such activity? The authors believe that there is a case for such usages, that they will, ultimately, be attractive and that the work described above is central to the realisation of that goal.

The proposed usages could also pose issues for regulation. The usages are essentially communication activities so we might expect them to be subject to communication regulation. But the experiences could be funded by advertising, and advertising regulation is usually delivered through a media regulator. The

[23] Digital Living Network Alliance: www.dlna.org

usages would also raise issues on security and privacy; users might well expect that experiences shared between friends using the system would be as private as a similar interaction in "real life". Such assurances could only be delivered (given that the experience has been recorded as an essential part of its creation) through some form of regulation.

A further complication could arise when communication occurred between different national jurisdictions of regulation. For example alcohol or tobacco based product placement may be allowable in some countries but not in others. A framework would need to be in place to deal with such occurrences

The Audio Visual Media Services Directive (AVMSD) seeks to define a common framework for European regulation on moving image media, including a framework to ensure the protection of minors, on advertising rules and on support for those with hearing or eyesight difficulties. However it is not clear yet whether the usages described would fall under the jurisdiction of the AVDMS directive or whether they would be covered by communications regulation, which usually deals more with ensuring competitive market where significant market power cannot be abused. So whilst at this proposition stage, it is not clear how to address the arising regulatory issues, it seems likely that the regulation of such usages would be problematic when regulatory bodies are tightly focused on either communications or media.

We postulate that, with media and communications converging it seems sensible, indeed necessary, for regulatory bodies for media and communications to at least liaise closely with each or even for the regulatory bodies to merge in order to adequately regulate such activities.

8. CONCLUSIONS

The paper has argued that the television screen could be used to support significant new forms of social communication between groups of people in different households. The argument is supported by social science theory. We have developed a number of loosely sketched framing experiences that have subsequently been evaluated in a user-centred fashion through semi-structured qualitative interviews with 16 families across 4 European countries. The interviews uncovered valuable qualitative insights into people's behaviour and patterns of social communication. The following findings were thought to be significance for our research:

- Playful activity fulfilled an important social function for many (though not all) of the families interviewed.
- Many families cited the difficulty of using technology as a barrier to its adoption and domestication.
- The use of communication tools to help cope with caring situations was a recurring theme for our participants, and they perceived that better and richer communications could potentially support this role in emotionally trying times.
- Family members were not generally regular users of video communication in any form, citing technical deficiencies and dullness as reasons why it remained under used.
- Some interviewees did comment on the value they perceived video bringing; this included "*seeing* how people were", and gaining a sense of 'being with' someone through an always-on video feed from a webcam. One family cited improved turn taking in group-to-group communication as a further potential benefit.

- Many families used media they had created themselves (particular photographs) as a means of stimulating and supporting social communication, though most families were careful to share though private channels (such as email or private web spaces). The use of home video was limited and viewing of home videos was even more limited, with dullness being cited as a significant contributing factor.

The findings from these interviews, allied with understanding from social sciences about the fundamental needs of humans and on the role visual communication plays in the communications of feelings and attitudes, have been used to challenge and validate the focus of proposed technical work required to deliver a range of new applications that should help support social communication.

Four hypotheses have been generated about how to improve the nature of the framing experiences discussed in this paper if they are to succeed in encouraging social communication between groups:

- Engagement, entertainment and excitement should be brought into video based representations used in communications in order that it provides a quality similar to that of professionally-crafted TV programmes, which have been proven to be able to engage the viewer.
- Flexibility is required in the designing of framing experiences to allow users to communicate within and around the main focus of the framing experience (which may be a game).
- High quality, reliable audiovisual communications will encourage usage of the framing experiences. We propose that high-quality videoconferencing capabilities should be brought into the domestic environment and made to operate on high definition TVs over contended broadband networks.
- The twin and often conflicting masters of control and ease of use must be accommodated in framing experiences; this should be achieved through rigorous user centred–design processes.

Finally a number of technology components were highlighted that could be used in order to test the hypotheses generated above. These included:

- automatic orchestration of the audio visual communications (like an automatic video mixer or director – this will deliver the visual excitement and entertainment required to keep the image interesting)
- multimedia interpretation (to automatically provide cues with which decisions on automatic orchestration can be made)
- multimedia composition and delivery developments (in order to dynamically composite on the screen the action and intentions of the automatic orchestrator – this will also help keep the image interesting)

The usages described are seen to have two significant implications on government policy and industry strategy.

- In terms of regulatory policy the convergence suggest that regulatory policy making bodies should also collaborate and merge to address the issues facing convergent industries more adequately

- In terms of industry strategy the usages suggest that greater effort should be placed in identifying, and then recommending, capabilities and standards for consumer equipment that will reliably support real time communications services constructed from disparate technology artefacts.

In summary an extensive and enlightening user centred design activity has been completed. The activity demonstrates how qualitative interviews at the early stages of a design process can highlight challenges that can be addressed through technology and identifies four hypotheses that will now become the subject of further experimentation and evaluation.

9. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2007-214793.

10. REFERENCES

[1] Alderfer, C.P., (1969) An Empirical test of a New theory of Human Needs *Organizational Behaviour and Human Performance* vol 4, pp142-175

[2] Aroyo, L., Bellekens, P., Bjorkman, M., Houben, G.J., Akkermans, P., and Kaptein, A. (2007). SenSee Framework for Personalized Access to TV Content. In *Proceedings of EuroITV*, pp. 156-165.

[3] Battelle, J., Packaged Goods Media vs. Conversational Media, *John Battelle's Searchblog, 5th December 2006*, http://battellemedia.com/archives/003160.php

[4] Blumler J. G. & E. Katz (1974): The Uses of Mass Communication. Newbury Park, CA: Sage

[5] Bocconi, S., Nack, F., and Hardman, L. Vox Populi: A Tool for Automatically Generating Video Documentaries, in *Proceedings of the Sixteenth ACM Conference on Hypertext and Hypermedia*, Salzburg, Austria, pp. 292–294, 2005.

[6] Boertjes, E., Klok J., and Schultz S. ConnecTV: Results of The Field Trial. Available online at http://soc.kuleuven.be/com/mediac/socialitv2/papers/ConnecTV_Results_of_the_Field_Trial.pdf, Proceedings of EuroITV Conference 2007, Amsterdam, The Netherlands, 2007.

[7] Cesar, P., Bulterman, D.C.A., and Soares, L.F.G. Human-Centered television: directions in interactive digital television research. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 4(4): 24, 2008a.

[8] Cesar, P., Bulterman, D.C.A., Geerts, D., Jansen, J., Knoche, H., and Seager, W. Enhancing social sharing of videos: fragment, annotate, enrich, and share. In *Proceedings of the ACM Conference on Multimedia*, pp. 11-20, 2008a.

[9] Chorianopoulos, K. and Lekakos, G. Social TV: Enhancing the shared experience with interactive TV. *International Journal of Human-Computer Interaction*, 24(2) pp. 113-120, 2008.

[10] Coppens T., Vanparijs F. and Handekyn K, "AmigoTV: A Social TV Experience Through Triple-Play Convergence", *Alcetel-Lucent white paper*, 2005.

[11] Davis, A., A Ready Market: Introducing H.264 SVC, *Wainhouse Research 2006*

[12] Ducheneaut, N., Moore, R. J., Oehlberg, L., Thornton, J. D., and Nickell, E. SocialTV: Designing for distributed, social

television viewing. *International Journal on Human Computer Interaction*, (24):2, 136-154, 2008.

[13] Forgas J (ed,. 2005) *Social Motivation: Conscious and Unconscious Processes Cambridge University Press*

[14] Del Galdo, G., Kuech, F., Kallinger, M., and Schultz-Amling, R. 2009. Efficient merging of multiple audio streams for spatial sound reproduction in Directional Audio Coding. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.19-24

[15] Han, I., Park, H., Choi, Y., Park, K., Four-way video conference and its session control based on distributed mini-MCU in home server, *Proc. IEEE International Conference on Consumer Electronics (ICCE '08)*, pp. 233-4

[16] Harboe, G., Massey, N., Metcalf, C., Wheatley, D., and Romano, G. The uses of social television, *ACM Computers in Entertainment*, 6(2): 8, 2008.

[17] Harmon, A., Grandma's on the Computer Screen, *New York Times, 27th November 2008*, p.A1

[18] Hemmeryckx-Deleersnijder B., and Thorne, J.M. Awareness and conversational context-sharing to enrich TV-based communication. *ACM Computers in Entertainment*, 6(1): 7, 2008.

[19] Kubey, R. and Csikszentmihalyi, M. Television and the Quality of Life: How Viewing Shapes Everyday Experiences. Lawrence Erlbaum, 1990

[20] Larsson, H., Lindstedt, I., Lowgren, I., Reimer, B., and Topgaard, R. From Time-Shift to Shape-Shift: Towards Nonlinear Production and Consumption of News. In the Proceedings of the EuroITV 2008 Conference, Salzburg, Austria, pp. 30–39, 2008.

[21] Lull, J. Family Communication Patterns and the Social Uses of Television. Communication Research, 7(3) pp.319-333, 1980.

[22] Metcalf, C., Harboe, G., Tullio, J., Massey, N., Romano, G., Huang, E.M., and Bentley F. Examining presence and lightweight messaging in a social television experience, *ACM Transactions on Multimedia Computing, Communications, and Applications*, 4(4): 27, 2008

[23] Masthoff, J. (2004). Group modeling: Selecting a sequence of television items to suit a group of viewers. User Modeling and User Adapted Interaction. 14, pp. 37-85.

[24] Maslow "A Theory of Human Motivation" A.H Maslow *Psychological review* 50, pp. 370-396, 1943

[25] Mateas, M., and Stern, A. Structuring Content in the Facade Interactive Drama Architecture. In Proceedings of the *First Annual Artificial Intelligence and Interactive Digital Entertainment Conference*, AAAI Press, New York, 2005.

[26] Mateas, M., Vanouse, P., and Domike, S. Generation of ideologically-based Historical Documentaries. In the Proceeding of *The Conference of the Association for the Advancement of Artificial Intelligence*, Austin, pp. 36–42, 2000.

[27] Mehrabian, A "Silent Messages: Implicit Communication of Emotions and Attitudes 2nd ed" pp. 75 -80, California, Wadsworth, 1981

[28] Nathan, M., Harrison, C., Yarosh, S., Terveen, L., Stead, L., and Amento, B. CollaboraTV: making television viewing social again. In *Proceedings of the International Conference*

on Designing Interactive User Experiences for TV and Video, pp. 85-94, 2008.

[29] Odlyzko, A "The Delusions of Net Neutrality" 36th Telecommunications Policy Research Conference 2008 Ricci, E., and Odobez, J.M. 2009. Learning Large Margin Likelihood for Realtime Head Pose Tracking. In *Proceedings of the IEEE International Conference on Image Processing* (ICIP).

[30] Rogers, E. M. 1995. Diffusion of innovations, Fourth edition. New York: The Free Press.

[31] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and Schooler, E., IETF Request For Contributions 3261 SIP: Session Initiation Protocol, June 2002, available from *http://www.ietf.org/rfc/rfc3261.txt*, 2002.

[32] Shamma, D.A., Bastea-Forte, M., Joubert, N., and Liu, Y. Enhancing online personal connections through the synchronized sharing of online video. In *Extended Abstracts on Human Factors in Computing Systems*, pp. 2931–2936, 2008.

[33] Tuomola, M. (dir), Saarinen, L.E., and Nuurminen, M.J., Accidental Lovers, Crucible Studio, Helsinki University of Art and Design and YLE, Finland Public Broadcasting Company, December 2006 – January 2007, 2006/2007

[34] UK Office National Statistics "Family Spending: 2007 edition" p4 & Table 4.2 pp68 -69,Palgrave Macmillan.

[35] Ursu, M.F., Kegel, I.C., Williams, D., Thomas, M., Mayer, H., Zsombori, V., Tuomola, M.L., Larsson, H. and Wyver, J. ShapeShifting TV: Interactive Screen Media Narratives, *ACM/Springer Multimedia Systems* 14 (2), pp. 115–132, 2008a.

[36] Ursu, M.F., Thomas, M., Kegel, I., Williams, D., Tuomola, M., Lindstedt, I., Wright, T., Leurdijk, A., Zsombori, V., Sussner, J., Maystream, U., and Hall, N. Interactive TV Narratives: Opportunities, Progress and Challenges, *ACM Transactions on Multimedia Computing, Communications and Applications*, 4(4): Article 25, pp. 25:1–25:39, 2008b.

[37] Zsombori, V., Ursu, M.F., Wyver, J,, Kegel, I., and Williams, D. ShapeShfting Documentary: A Golden Age. In the Proceedings of the EuroITV 2008 Conference, Salzburg, Austria, pp. 40–50, 2008.