

AN EMERGENT ROLE FOR TV IN SOCIAL COMMUNICATION

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

Joshua Meenowa
BT
Adastral Park
IPSWICH, IP5 3RE, UK
+44 (0)1473 643365
joshua.meenowa@bt.com

ABSTRACT

This paper discusses early results from the EU FP7 project Together Anywhere Together Anytime (TA2). TA2 is related to the use of television as a device which supports social interactions between groups. The paper introduces the idea of framing experiences, within which social communication can take place, and explains that social communication is a key motivator in a number of theories on human motivation. The paper reports how some ideas from framing experiences were evaluated through semi structured interviews with 16 families across 4 European countries. It discusses how some of the key issues and behaviours mentioned by users generate a number of hypotheses about how such communications could be improved.

Categories and Subject Descriptors

H.5.2 [User Interfaces] User-centred-design, evaluation/methodology. **H.5.3 [Group and Organisation Interfaces]** Asynchronous interaction, synchronous interaction. **J.4 [SOCIAL AND BEHAVIOURAL SCIENCES]** [Psychology] [Sociology]. **I.5.2 [Design Methodology]** Feature evaluation and selection

General Terms

Design, Experimentation, Human Factors

Keywords

Social communications, orchestration

1. INTRODUCTION

The television's affectionate place in the mind of society is as a social family activity [18][20], capable of binding significant fractions of a nation's audience to a shared experience. The first

generation of interactive television research has imported a number of successful web-based models without taking into account that television watching is normally a shared activity. Media consumption on a computer is, in contrast, usually an individual experience; personalized television with obtrusive overlays over the content is an example of such tendency.

Fortunately, over the last couple of years, we have seen an increased interest on interactive television as a shared experience. For example we can find some solutions that consider group modelling for content recommendation. [2][22] At the same time, progress is being made on television experiences that can be shared in real time between households. [15][17][10]

The paper reports on work completed in the EU FP7 project Together Anywhere Together Anytime (TA2) that define and validate a range of new social, communication based TV experiences that are intended to resonate with the view of television as a social and shared experience by facilitating social interaction within and between households.

The motivation and objectives section describes why, from a high level psychological standpoint, it is believed the posited approach is likely to be successful and goes on to introduce the methodology of some research (of which this work forms a part) that attempts to design, build and validate a number of framing experiences that attempt to support social communication and interaction through the television screen.

The state of the art section describes relevant research in social and interactive TV, communication in screen based games, video communications and composition in interactive narratives, and media sharing.

The key results for this paper are based on the perceptions of users and are reported in the 'family interviews' section. This reports qualitative results of research into social communication behaviours based on interviews with sixteen families in four European countries.

The technology section responds to the results of the family interviews, developing a number of hypotheses on how technology can be used to address issues raised in the interviews.

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

EuroITV'09, June 3-5, 2009, Leuven, Belgium.

Copyright 2009 ACM 978-1-60558-340-2/09/06...\$5.00

2. MOTIVATION AND OBJECTIVES

Whilst improving social communication may be an inherently valuable goal, this research is driven by a simple profit motive. Historically the provision of communications, as traditional telephony, or more recently mobile phones or indeed the Internet can be very profitable. The ultimate goal of this work is to explore new forms of rich social communications that will bridge the benefits of tools for social group communications such as Facebook™¹ and Twitter™² with the richness and transparency of video conferencing.

This work supposes that people are motivated to conduct social communications. 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 [13]; what does it matter whether actions are goal directed or sub conscious? Provided there is a motivation within people to be sociable and to communicate then the foundations of this work are supported.

We note that the key models of Maslow, the hierarchy of needs [21], the existence, relatedness and growth model (ERG) [1] and, in a more media-specific vein the uses and gratification (U&G) model [4] all have, as central to their models, the need for social communications. Maslow speaks of love needs, Alderfer of relatedness and Blumler and Katz speak of personal relationships, all of which will be nurtured by social communications.

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 [31] 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 [31] (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 expenditure, 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. [25]

ICT systems are, particularly between domestic settings, 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

¹ Facebook™: <http://www.facebook.com>

² Twitter™: <http://www.twitter.com>

assessed within some kind of activity designed to allow social communications 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 therefore seek to design and build a number of such framing experiences, and to evaluate their usage. The 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.

The objective of the work described in this paper is to develop a number of hypotheses about the potential impact of certain technological developments and to design framing experiences in which we can hope to test such hypotheses.

3. METHOD

Television still plays a central social role in households. It is posited that it should be possible to further exploit this role in the enhancement of social communications 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.

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 early 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 'framing experiences' were drafted, each designed to offer people the opportunity to nurture relationship with friends or 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 allowing “teams” from different households to explore a game world and for that experience and the game play to be dependent upon real time voice and video communications 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 or football tricks or skateboard tricks) and to invite their friends to copy and learn from them and to practice 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 audio and video material they have shot themselves and on material shot by others.

4. STATE OF THE ART

This section introduces the state of the art in a number of relevant technology areas: social and interactive TV, communication in screen based games, video communication and composition, interactive narratives and media sharing.

4.1 Social and interactive TV

The research field of interactive digital television is being transformed into a study of human-centred television [7], in which television viewers become active nodes 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, there is currently 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 node, the television viewer might want to communicate with others while watching [9][12], to leave notes and comments for friends at specific moments of a television show [27], and to share enriched fragments of multimedia content with others [8]. For example, TV-based services like Alcatel-Lucent’s Amigo TV [10] allow users to watch broadcast TV together (when apart) and to augment their watching experience with voice chat, messaging and the use to emoticons to create a social shared TV viewing experience. The Social Television project [26] 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 [29] 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 the direct communication link (e.g., text chat or audio chat) as a meta-activity, rarely related to the actual 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 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.

4.2 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. Note that all layers are not 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, discussing strategies, memorable moments, et cetera. 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*³, *Teamspeak*⁴ or *X-box Live*⁵. In this phase, the content of the communication is mainly focused on the gameplay itself; giving orders and heads up, discussing strategy, et cetera. 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. A case could be made that the success of *Counterstrike*⁶ stems partly from the inability of gamers to respawn immediately when their in game characters die; waiting to be respawned provides the opportunity to discuss the game online amongst themselves, reinforcing 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

³ Ventrilo Client and Server (1999-2007) Flagship Industries Inc

⁴ Teamspeak (2008) TeamSpeak Systems GmbH

⁵ Xbox Live (2008) Microsoft Corporation

⁶ 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 that the users will stay users of said system for a greater period of time. *Wii Speak*⁷ seems to strive in this exact direction as it facilitates many different layers of communication both 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.

4.3 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*⁸ and *Telepresence*⁹. These almost always require 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. During the last year, manufacturers have begun to announce¹⁰ 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. [14] 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 [11] 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 [28] and IMS¹¹

[SIPK 2008] 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 [16] that 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)¹² are building interoperability standards for consumer equipment [DLNA 2007] 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 standard¹³, 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 SMIL¹⁴ 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.

4.4 Interactive TV narratives

Interactive TV narrativity is a subset of interactive television in which the (active) viewer can influence the programmes (the stories) that they receive. The relationships between this new form of television and the framing experiences is evident. First, social communication is normally framed as a narrative. Interactive narratives allow end users to "converse" with the narration (with a virtual storyteller). Our framing experiences require this kind of communication, but take it further: interactive TV narratives, or interactive TV programmes, are authored by experts and subsequently delivered in an interactive manner to the active viewers; the framing experiences referred to here do indeed require interactive narratives, but they cannot be 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* [23] and *Accidental Lovers* [30]; *Terminal Time* [24], *Vox Populi* [5] and *A Golden Age* [34]; and

¹² DLNA, Overview and Vision White Paper 2007, available from <http://www.dlna.org>, 2007

¹³ UPnP, AV Architecture:1 (Approved Design Document), 25 June 2002, available from <http://www.upnp.org/specs/av/>, 2002.

¹⁴ SMIL, Activity Statement for W3C Synchronised Multimedia Working Group, available from <http://www.w3.org/AudioVideo/Activity.html>, 2008.

⁷ WiiSpeak (2008) Nintendo

⁸ Hewlett-Packard Halo Collaboration Studio

⁹ Cisco TelePresence conferencing suites

¹⁰ <http://www.crn.com/networking/208801088>

¹¹ SIPK, 3GPP IMS Specification List, available at http://www.sipknowledge.com/IMS_Specs.htm, 2008.

My News And Sports My Way [19]. 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 [33].

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 [32], represents a good paradigm for this.

4.5 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*, *Jumpcut* 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' [3].

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 'remix' stories from multiple uploaded fragments.

A subset of user-generated online video sites (*Jumpcut*, for example) 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. 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 family interviews.

5.1 Method

Sixteen families across 4 countries (UK, Sweden, Netherlands and Germany). 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 concepts that TA2 is considering developing as described in section 4.

Interviewees were in family groups, with children aged between about 6 and 25. The households tended to be of higher than average income and included early adopters as well as "middle majority" and some laggards in terms of technology adoption. We explicitly sought households with long standing, deep bonds with another household and which yearn to be closer together but seldom found the chance to meet up anymore.

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 different types of social interactions: play and sharing emotions, currently take place and discuss the reactions of our participants to our ideas relevant to these fields. We then look at the place audiovisual communications take 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.

5.2 Results

The qualitative approach is not statistically representative, nevertheless the interviews are valuable as our investigations and solutions are at the very early stages and are still taking shape. Furthermore, given the limiting costs of interviews, the qualitative study we report here best fits the stated goal of informing our designs.

5.2.1 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. In particular, interviewees in Sweden 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, such as Wii games which were very popular during meet ups and parties. Closer to 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 such non-task oriented intrapersonal interplay behaviour such as teasing, discussions, interruptions and rules and strategy evolution.

Some household members also said 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.

5.2.2 *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 Skype¹⁵ 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 Twitter-ing 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 Bebo¹⁶ or Facebook 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.

5.2.3 *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 webcams 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 VoIP 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.

5.2.4 *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.

5.2.5 *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 doing so: 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.

¹⁵ Skype™: <http://www.skype.com>

¹⁶ Bebo™: <http://www.bebo.com>

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.

5.2.6 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 Picasa¹⁷ or on disc or on Facebook, enabling friends and family to stay in touch with each others lives.

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.

5.2.6.1 Control of Shared Media

All participating parents, if they shared media, would do so via communication methods they perceived as private: the so called private Picasa 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 programmes, 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.

5.2.7 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 iPlayer¹⁸, 4oD¹⁹, 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 [6]. 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.

5.2.8 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.

6. TECHNOLOGY

The results of the family interviews described above support a number of hypotheses about how current technology can be extended beyond the state of the art in order to meet future users' needs for social, communication-based TV experiences.

6.1 Hypotheses

6.1.1 Excitement and Entertainment

While several families appreciated the value of quality non-verbal communication in aspects such as providing 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 be shared by the groups. Communication between these groups of people should be natural and require no active intervention. It should also provide a quality similar to that of professionally-crafted TV programmes, which

¹⁷ Picasa™: <http://picasaweb.google.com>

¹⁸ BBC© iPlayer™: <http://www.bbc.co.uk/iplayer/>

¹⁹ 4oD™: <http://www.channel4.com/4od/index.html>

have been proven to be able to engage the viewers in their space as if the space were real.

6.1.2 Flexibility

As described above, our family interviews discussed playful social activities between family members and introduced several framing experience concepts devised by the TA2 project. The families' reactions suggest that framing experiences should allow flexibility for interpersonal interactions which are not a part of the game itself.

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 should be embodied in audiovisual communication.

6.1.3 High-quality, reliable audiovisual communications

When the interviewees could envision technology enabling games to be played across distance, they highlighted their concerns about the reliability of our proposed technology, while our research has been concerned with the experience of the interaction

When using webcams, some participants 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 the game application, is combined with live audio and video streams.

6.1.4 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.

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

- 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.
- 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.

- 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 [32] is focused on creative professionals. These existing tools will be simplified to the extent they can be used by groups of families and friends.
- 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.
- The system must be able to adapt its user interfaces to the accessibility needs of diverse individuals simultaneously.

6.2 Technology Components

The TA2 project proposes to test these hypotheses through the incorporation of a number of technology components. These are:

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

To introduce the role of *interaction orchestration*, consider the example of a live TV programme presenting a main event that happens in different physical locations, such as "Red Nose Day" in the UK. The director has a number of views of the live events, through different cameras and microphones, but also has access to an archive of previously recorded material. He makes decisions on how to combine information from all these sources to best recount the main event in the broadcast. *Interaction orchestration* has a similar function, but in the context of the communication between the friends and family interacting in different scenarios developed in the project. 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 directing*. Continuing the metaphor, orchestration refers to all the decisions that the cameramen, the director and possibly the 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.

Media interpretation refers to the extraction of textual information possibly accompanied by audio and visual objects from captured media objects (and possibly ambient devices). At this early stage of TA2, two distinct challenges are anticipated: the extraction of audio and visual cues to enhance and capture the experience, and the detection of events and trends using these cues. 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.

TA2 software will have to be capable of dynamically combining live streamed content (for example from cameras) with pre-compiled compositions, creating a corresponding challenge in realising the media experience continuously being computed and

described for each participant. A number of unique problems arise in both the audio and video domains.

For example, the project seeks to advance the state of the art throughout the end-to-end audio subsystem by integrating tools for echo compensation, 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 bitstream mixing inside the codec domain will be developed which reduce complexity and delay, and increase quality of multipoint connections. New frame loss concealment algorithms, which make use of codec signal representations, will also be deployed to minimise delay and preserve audio quality.

7. CONCLUSIONS

The paper has argued, with support from social science, that the television could be used to support significant new forms of social communication between groups in different households. The idea of framing experiences was introduced and results from semi structured qualitative interviews with 16 families across 4 European countries, discussing exemplar framing experiences are reported. A number of significant qualitative insights into people's behaviour and patterns of social communication were identified. These included:

- Playful activity, loosely described as games, was an important social activity 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 which better and richer communications were sometimes seen as a useful (or potentially useful) support in difficult 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 through 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.

Interpreting these findings with an understanding from social sciences about the fundamental needs of humans and about the role visual communication plays in the communications of feelings and attitudes, has helped define the focus of proposed technical work required to deliver a range of new applications designed to help support social communications.

Four hypotheses have been generated about improving the nature of the framing experiences discussed in this paper:

- Entertainment and excitement should be brought into video based representations used in communications for it to engage the viewer in the same manner as professionally-crafted TV programmes.
- Flexibility is required in the designing of framing experiences to allow users to communicate within and around the main focus of the framing experience e.g. 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 matters of control and ease of use must be accommodated in framing experiences- this should be achieved through rigorous user centred design.

Finally, a number of technology components were highlighted to test the hypotheses generated above. These included:

- automatic orchestration of the audio visual communications,
- multimedia interpretation to automatically provide cues to prompt automatic orchestration,
- multimedia composition and delivery developments.

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.

8. 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.

9. 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 the European Conference on Interactive Television*, 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/sociality2/papers/Connec>

- [TV_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, 2008b.
- [9] Chorianopoulos, K. and Lekakos, G. Social TV: Enhancing the shared experience with interactive TV. *International Journal of Human-Computer Interaction*, 24(2):113-120, 2008.
- [10] Coppens T., Vanparijs F. and Handekyn K, "AmigoTV: A Social-TV Experience Through Triple-Play Convergence", *Alcatel-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 et al., (ed., 2005) *Social Motivation: Conscious and Unconscious Processes Cambridge University Press*
- [14] 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
- [15] Harboe, G., Massey, N., Metcalf, C., Wheatley, D., and Romano, G. The uses of social television, *ACM Computers in Entertainment*, 6(2): 8, 2008.
- [16] Harmon, A., Grandma's on the Computer Screen, *New York Times*, 27th November 2008, p.A1
- [17] 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.
- [18] Kubey, R. and Csikszentmihalyi, M. *Television and the Quality of Life: How Viewing Shapes Everyday Experiences*. Lawrence Erlbaum. 1990
- [19] 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.
- [20] Lull, J. Family Communication Patterns and the Social Uses of Television. *Communication Research*, 7(3):319-333, 1980.
- [21] Maslow, A.H. "A Theory of Human Motivation", *Psychological review* 50, 370-396, 1943
- [22] 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.
- [23] 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.
- [24] 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, 36-42, 2000.
- [25] Mehrabian, A "Silent Messages: Implicit Communication of Emotions and Attitudes 2nd ed." pp75 -80, 1981, California, Wadsworth
- [26] 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
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] UK Office National Statistics "Family Spending: 2007 edition" p4 & Table 4.2 pp68 -69, Palgrave Macmillan.
- [32] 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.
- [33] 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.
- [34] Zsombori, V., Ursu, M.F., Wyver, J., Kegel, I., and Williams, D. ShapeShifting Documentary: A Golden Age. In the Proceedings of the EuroITV 2008 Conference, Salzburg, Austria, pp. 40-50, 2008.