School of Electronic
Engineering and
Computer Science

Media and Arts
Technology Programme
Project Report 2011

Conversational Annotation via Social TV

Saul Albert
Student Nr. 100464131
August 2011

Contents

1 Acknowledgements

Thanks to my two supervisors Andy Gower at BT, and Pat Healey at Queen Mary. Thanks also to Richard Kelly, Toby Harris, and *The People Speak*. This research was funded by the Digital Economy programme through the Media and Arts Technology Doctoral Training Centre at Queen Mary, University of London.

2 Abstract

What are the relationships between content and communication? Their often-cited “convergence” on the Internet, and the overwhelming availability and diversity of networked media has seen TV broadcasters adopt paradigms of user profiling and data mining from Social Networking sites in order to provide users with “Social TV” services that broker their interactions and consumption patterns in order to offer content recommendations.

This research project questions this approach by exploring the differences between “content” as understood by broadcasters and advertisers, and the way it is demonstrably used in viewers interpersonal communication.

The approach here is to analyse formal, broadcaster-supplied content ontologies, alongside ontologies implied by data derived from an existing “sociable” behaviour of TV viewers: the heckling, interjection and conversation that characterises the convergence of content and communication in the living room.

The “Heckle” tool is developed and tested with participants in a series of screenings of an episode of the TV show *Doctor Who*, for which highly granular broadcaster-supplied metadata are available.

A detailed comparison of the two data sources provides the basis for analytical results that support existing evidence about thematic (but not lexical) alignment, and the common grounding of viewer interaction in media content. New evidence is presented to substantiate and critique claims about what factors determine the semantics of content annotation, and new insights and research directions are gained through a close study of “conversational annotation”.

Conversational annotation is then proposed as a complimentary representation of media objects in their social context, one that is qualitatively different from formal content metadata, and suggests different attitudes and practical approaches towards brokering the convergence of content and communications.

A PLAY IN A LONDON INN YARD, IN THE TIME OF QUEEN ELIZABETH.

From Thornbury's *Old and New London*, Cassell & Co, 1881.

Figure 1: The raucous audience of a play in a London inn yard, in the time of Queen Elizabeth, from Thornbury's Old and New London, Cassell & Co, 1881.

3 Introduction

The convergence of content and communications via the Internet has created the opportunity for new kinds of television services, where information can flow from broadcasters to consumers and vice versa (Chorianopoulos, 2007), as well as between viewers themselves and other agents, such as Internet-enabled televisions, mobile devices, broadcasters, content producers and third party services such as Social Networking platforms (Klym & Montpetit, 2008), (Bernhaupt et al., 2008), (Harboe, 2009).

"Social TV", is becoming the accepted catch-all term for services that support or extend the "sociable" aspects of television viewing (Harboe, 2009), (Coppens & Trappeniers, 2004), (Oehlberg et al., 2006), and Social TV services are now being enthusiastically deployed by technology entrepreneurs,

broadcasters and content providers (Cesar & Geerts, 2011).

However, despite the findings of the few Social TV research projects (Shamma et al., 2007), (Nathan et al., 2008), (Fagá Jr et al., 2010) that have explored the complexity of TV-centric social behaviour, many Social TV systems seem to derive their model of "sociability" from the Social Networking platforms with which they are becoming increasingly integrated (Aroyo et al., 2009), (Schopman et al., 2010).

For example, many specialised Social TV systems such as GetGlue[1] or Tunerfish[2] "reward" viewers for "checking in" to certain programmes, mimicking the location-based game-dynamics of spatial marketing service Foursquare[3]. By brokering interactions between users and TV content (Cremonesi & Turrin, 2010), these services gather marketing data that is used to power search and recommendation engines that ease the discovery and use of TV content (Melville et al., 2002), (Yu et al., 2006). By being more findable, a TV show becomes more valuable to broadcasters who sell advertising around it, rights-holders who gain royalties from it, and presumably, viewers who want to watch it (Van Aart et al., 2009).

However, by concentrating on enabling interaction between remote viewers, and adopting paradigms of sociability developed for web-based or mobile Social Networks, Social TV systems can be seen to underplay the complex interactional scenario of co-present TV viewing (Svensson & Sokoler, 2008), in which the TV screen becomes a contextual cue for conversation between people watching together: a "ticket to talk" (Sacks, 1992).

Seeing Television as a viewing event "embedded" in social activities, rather than a distinct technological or narrative media object (Harvey, 1990), suggests that it is "not solely about viewer and viewed, but also viewer and viewer" (Butsch, 2003, p.19). This understanding of television questions the assump-

[1] http://www.getglue.com

[2] http://www.tunerfish.com

[3] Foursquare: http://www.foursquare.com is an application for GPS-enabled mobile "smartphones" that invites users to visit geolocated "places", shops, restaurants or tourist sites, and "check in", earning symbolic rewards such as "badges" and "mayorships" for regular attendance, and discretionary discounts from proprietors.

tions in much Social TV research that the integration of communications should avoid “distracting” viewers from content (Geerts, 2006), (Weisz & Kiesler, 2008), (Cesar & Geerts, 2011).

Instead, like the rowdy audiences of early Nickelodeons and silent film (Hansen, 1994), or even earlier, the bawdy audiences of Elizabethan theatre (Brown, 2002)[4], the promise of Social TV highlighted in early research (Oehlberg et al., 2006) is that its services, and the ways in which it makes TV content available can be shaped by interactions between viewers themselves, and by the ways they deploy TV content as a part of their communications (Shamma et al., 2007).

3.1 Research Question

What is the difference between the ways broadcasters understand TV content, and the ways it is used in social communication between viewers?

Within this central question are several more embedded: what is meant by the words“television”, “social” and “content” in this context? And what pragmatic approaches can be taken to evaluate their differences?

Re-phrasing the research question to address this last point first: what are the differences between the formal ontologies used by broadcasters to describe their content, and those implied by transcripts demonstrating ways in which that content is used socially?

This research project uses a critical reading of Social TV as a research context to explore the differences between “content” and “communication”, and to start looking at ways to exploit convergence to understand, describe and organise media content based on its communicative uses.

[4]See figure 1

3.2 Summary

A review of Social TV research establishes methodologically grounded understandings of the terms “television”, and “social” and ends with a brief overview of the different technical, theoretical and pragmatic approaches to describing “content” in specific situations of use.

The “Heckle tool”: a system designed to elicit and capture the interjections, comments and conversations between TV viewers is described, along with a report of its use in two screenings designed to capture the interaction of groups of viewers watching the same episode of the TV show *Doctor Who*, for which detailed broadcaster-supplied metadata were available.

Methods based on qualitative observation and simple pragmatic measures are deployed to gather data for a comparative study of the formal ontologies of broadcaster-supplied content metadata and the implied ontologies of the practical use of media content in the conversations “heckled” while watching together.

Results demonstrate an overlap in conceptual grounding between the two data sets, although study of the relative alignment of the data to different conceptual grounds (the video, or the interactional context) illustrates the different contingencies of their production processes.

A close study of the interactional data determines that it can be seen and analysed as conversational, and suggests that the way people watching together negotiate topics of conversation while shifting between context and content orientation suggests different approaches for how content is organised and made available.

4 Literature Review

4.1 Purpose

This review concentrates on developing a methodologically grounded understanding of what is meant by the terms “television”, and “social”, and how researchers have brought them together in the context of Social TV to perform speculative field tests and analyses. A brief overview of general literature relating to formal ontologies and semantics is explored to develop a rationale for conducting a comparison between formal media metadata and informal conversational transcripts.

4.2 Social TV

4.2.1 Television

“Social TV” as a research area refers to systems that support any social practices associated with TV viewing, including talking about, watching or recommending TV shows (Oehlberg et al., 2006), (Harboe et al., 2008a). Many researchers also use the term more narrowly (Harboe, 2009) to refer to specific technologies, devices and communication modalities intended to enable remote groups of viewers to evoke the experience of “sharing the couch” (Harboe et al., 2008a), while watching TV together (Cesar & Chorianopoulos, 2007), (Schatz et al., 2007), (Klym & Montpetit, 2008), (Luyten et al., 2006), (Geerts & De Grooff, 2009).

However very little Social TV research has reported using ethnographic observational study of the existing behaviours of television viewers as a starting point with a few notable examples (Oehlberg et al., 2006), (Bernhaupt et al., 2008). Instead most studies have developed prototypes for various kinds of Social TV systems, for example, communication-oriented Social TV systems to compare the affects of integrating different modalities such as open audio channels (Coppens & Trappeniers, 2004), (Colaco & Kim, 2010), (Harboe et al.,

2007), and text chat (Geerts, 2006), (Weisz et al., 2007), (Tullio et al., 2008), alongside TV viewing, and then observed use of those prototypes. This approach raises a question about what part of the experience of watching TV "Social TV" research is intended to evoke.

This is part of a wider question about how to understand Television in various branches of research. Hartley (1999) treats it variously as a "socio-personal" phenomenon, a domestic environment, a formal object of academic study examining mass society, television as a text, as an audience, or as a pedagogical tool. Zillmann & Bryant (1985) argue that television is most often defined by specific programme content, the dominant typology being the "TV Show". They ascribe this to research being driven by industry concern with viewer ratings, and to the relative ease and manageability of segmenting data into "theoretically relevant content categories". Critiquing the devices of measurement used in industry and research, specifically the Neilsen group's Set Meter and viewer diaries systems (Buzzard, 2002), Ang (1996) describes this focus on TV content and naive viewer measurement as the "black box model" of viewer behaviour, revealing the television industry's "calculated ignorance" about the "tactics by which consumers constantly subvert predetermined and imposed conceptions of watching television" (Ang, 1996, p.55).

The technologies of Social TV may enable more sophisticated audience measurement, including a granular, "personalised" model of "relevant content categories" (Yew et al., 2011), as well as providing diverse measures of viewer activity and communication from remote control use and volume changes to social network relationships (Aubert & Prié, 2005). However, a review of the literature suggests no solid technical nor ethnographic grounding for television itself as an observable viewer activity that could be used to pry open Ang's "black box".

4.2.2 "The Social"

Sociological literature from the pre-consumer Internet era documenting observation of the social behaviour of TV viewers (Lull, 1980), (Silverstone & Morley, 1990) is often cited in Social TV research (Weisz et al., 2007), (Baca, 2008), (Oehlberg et al., 2006), (Geerts & De Grooff, 2009) to claim that "[s]ince its inception television watching has been a social activity" (Cesar & Geerts, 2011, p.348), and to counter prior assumptions and contemporaneous assertions, notably by Robert Putnam in the mid-90's, that television-watching is an inherently anti-social, isolating activity that diminishes the "social capital" of civic life (Putnam, 1995), (Campbell et al., 1999), displacing what his thesis characterises as "social time" that people would otherwise spend in public spaces.

Putnam's more recent popular book "Bowling Alone" Putnam (2001) updates his thesis for the Internet era. Ironically, this book is also cited by many Social TV researchers (Cesar & Chorianopoulos, 2007), (Nathan et al., 2008), (Barkhuus, 2009), this time to substantiate the idea that "traditional joint television viewing" (Oehlberg et al., 2006), (now re-cast as an exemplary social activity), is under threat from the increasing use of laptops, media-enabled mobile phones and personalised or "catch-up" TV services restricting opportunities for co-present TV viewing, thereby diminishing the "water cooler effect" (Putnam, 2001) of prompting socialising about widely watched TV shows[5]. This argument is then used to introduce Social TV as a set of technologies and approaches that "may help to counteract the tendency of TV audience's fragmentation" (Abreu et al., 2002, p.3), by "making TV social again" (Nathan et al., 2008).

In an aside to their paper on the future of Social TV, Klym & Montpetit (2008) note a further irony: that Putnam's concept of "the social", which refers more to civic engagement in public spaces, than socialising in domestic spaces is

[5]He also mentions other factors such as urban sprawl, contemporary lifestyles, and the increase in the number of televisions in the house. Putnam's concern mirrors the anxiety of the television and media industries with the loss of control over what people watch, or 'audience fragmentation' (Cisco, 2010), (Goldmedia, 2010).

arguably better served by the convergence of Social TV systems and "sociable media" (a reference to Social Networking sites such as Facebook[6], Twitter[7] and other relatively "public" channels of communication) than by joint television viewing. However, the "publicness" of these networks, which Pold & Andersen (2011) have described as as "log-in spaces", seems debatable. Klym & Montpetit (2008), deferring to the assumed "sociability" of Social Networking (a common claim in much of the literature), rather than the social "experience of watching TV" calls into question Social TV's concept of "the social".

4.2.3 Watching Together

Ito and Okabe extrapolate their concept of the "technosocial situation" from a critical review (Meyrowitz, 1985) of Goffman's theories of "social situation" (Goffman, 1966), in which Meyrowitz cites television as a prime example of how electronic media transect and can reformulate the ways in which Goffman saw social practices as embedded in and contingent on particular social situations. Retaining Goffman's commitment to observing the particular, Ito and Okabe propose the "technosocial situation" as "a way of incorporating the insights of situationist theory into a framework that takes into account technologically mediated social orders" (Ito & Okabe, 2005, p.5).

In their critical review of Mobile TV literature, Harper et al. (2006) commend concept of the "technosocial situation" as a way to describe a new, flexible understanding of what Mobile TV might be: "[M]obile phone TV would not be TV shrunk down. But at the same time one would not expect mobile phone TV to be entirely new - it needs to evolve" (Harper et al., 2006, p:82). This is one of the key uses of Social TV research. As Shamma et al. note when questioning research methods that, in themselves, modify people's relationship with new media tools: "[i]deally, a virtuous circle forms: new tools make possible new ways of using media, generating new kinds of information about the social

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

functions of media content, which can in turn be fed back into the design of new systems." (Shamma et al., 2007, p.276).

Although its conceptions of "The Social" and "Television" seem ungrounded, Social TV here is seen as a particular technosocial situation, which through its pragmatic imperatives has developed and tested many transient forms of social and technological engagement with television, from which grounded methods, observations and analytical approaches can be derived.

The concept of the "Communication Space" (Healey et al., 2007) provides a complimentary approach to grounding a workable understanding of "sociability". In their paper, Healey et al. (2007) extrapolate on the work of Harrison (1996), in distinguishing between "place" and "space" as qualitatively different experiences and understandings of location, and a Heideggerian reading of interpersonal spatiality ("Being-with") to describe the concept of "Communication Space" as a space constituted by the "nearness and farness" (Healey et al., 2007, p.172) of others. The Communication Space functions here as a reminder of the primacy of interpersonal communication in repeatedly constituting and reconstituting "sociability", and the requirement to look at what is structuring the communication. The methods of Conversation Analysis (CA) (Sacks, 1995), which Healey et al. apply to the "Communication Space" of a text-chat environment[8] suggest that even without a definitive set of criteria for "sociability", the amenability of a corpus of interactional data to analysis by those methods can be seen as an indicator that the context in which the data was gathered functions, at least in part, as a "Communication Space".

Here, the "technosocial situation", and a cultural reading of Television and Social TV research[9] provide a workable grounding for what is meant by TV as a context for interpersonal interaction between viewers, in relation to a shared orientation to video "content". This research assesses the feasibility of performing CA on communication logs from group "heckling" at the screening of a

[8]The "Walford" Multi-User Dungeon (MUD), a text-based virtual reality in which users can construct and interact using (or transecting) spatial metaphors, creating what Healey et al. describe as a "rich communicative ecology" for their analysis.

[9]See appendix A.

video, and picking out the communicative tropes on which CA relies[10] from the resulting data in place of finding, and then testing against any formal metric for "sociability".

4.3 Content

4.3.1 Multimedia Metadata

Multimedia metadata is information about a media object, traditionally including both "low level" perceptual and technical characteristics such as dominant colours or spatio-temporal structure, and "high level" semantics relating to human interpretation (Brunelli et al., 1996). The cost, complexity and inconvenience of manually creating media metadata (Stamou et al., 2006), and the lack of support for the extraction and annotation of content metadata in widely adopted standards for multimedia manipulation and transmission such as MPEG-4[11] (Koenen, 2002) and MPEG-7 (Martínez, 2004) motivates extensive research in automated analysis and annotation systems (Petridis et al., 2006). However, a recent research survey found "the state of the art in computational perception is still somewhat limited to producing mostly low, signal-level metadata and some higher level metadata in constrained contexts." (Diakopoulos, 2009, p.12)

Although some "low level" features can be extrapolated to infer more complex semantics with a degree of reliability (Smeulders et al., 2000), newer metadata standards that adopt Semantic Web principles[12] for representing "high level" metadata cannot easily be incorporated into standard media formats (Van Ossenbruggen et al., 2004). Basic issues of syntactic incompati-

[10] Such as turn-taking, sequencing of adjacency pairs, and the crucial function of repair (the efforts made to understand and be understood) in providing evidence of inter-subjective communication (Schegloff, 1992).

[11] The Moving Picture Experts Group (MPEG) is the name of a family of standards for used for coding audio-visual information.

[12] Semantic Web principles here refer to the methods of Knowledge Representation advocated by Tim Berners-Lee including the development of machine-traversable ontologies, defining constrained vocabularies for storing and exchanging knowledge and the formal relationships between domain-specific "common sense" statements from which automated inferences can be drawn (Berners-Lee et al., 2001).

bility in multimedia metadata such as those between hierarchical, monolithic, domain specific XML[13] formats in MPEG-7, and unstructured, modular, and generalized RDF[14] tools in many Semantic Web technologies (Hunter, 2001) can be seen as symptomatic of their variable provenance and connections to different approaches to multimedia metadata. When defining the metadata standard, the designers of MPEG-7 had to balance the views of the various Knowledge Representation (KR) communities[15], pressing for the inclusion of "high-level" semantics, against those of the Signal Processing community who wanted to "standardise only low-level representations of content features and feature-detection algorithms" (Nack et al., 2005).

4.3.2 Media Semantics

An alternative method of representing media content such as a TV show, incorporating so-called "higher layers" (Tuffield et al., 2006) of semantics based on human interpretation of media content is proposed by the BBC Stories ontology[16]. Developed by Michael Jewell, Paul Riessen and Toby Harris in the context of a BBC research project (Harris, 2010), the Stories ontology is a specialised derivative of the OntoMedia ontology[17] (Lawrence et al., 2005), extending FOAF[18], Event[19] and Timeline[20] ontologies to enable machine-readable, traversable descriptions of detailed narrative elements such as characters, actions, scenes, places and plot developments (see figure 2).

[13] http://www.w3.org/XML/
[14] Resource Description Framework: http://www.xul.fr/en-xml-rdf.html
[15] Including researchers in the Digital Library (DL), Knowledge Representation (KR) and Multimedia for Artificial Intelligence (MM-AI) communities.
[16] http://www.contextus.net/stories/
[17] http://www.contextus.net/ontomedia
[18] http://xmlns.com/foaf/spec/
[19] http://motools.sourceforge.net/event/event.html
[20] http://motools.sourceforge.net/timeline/timeline.html

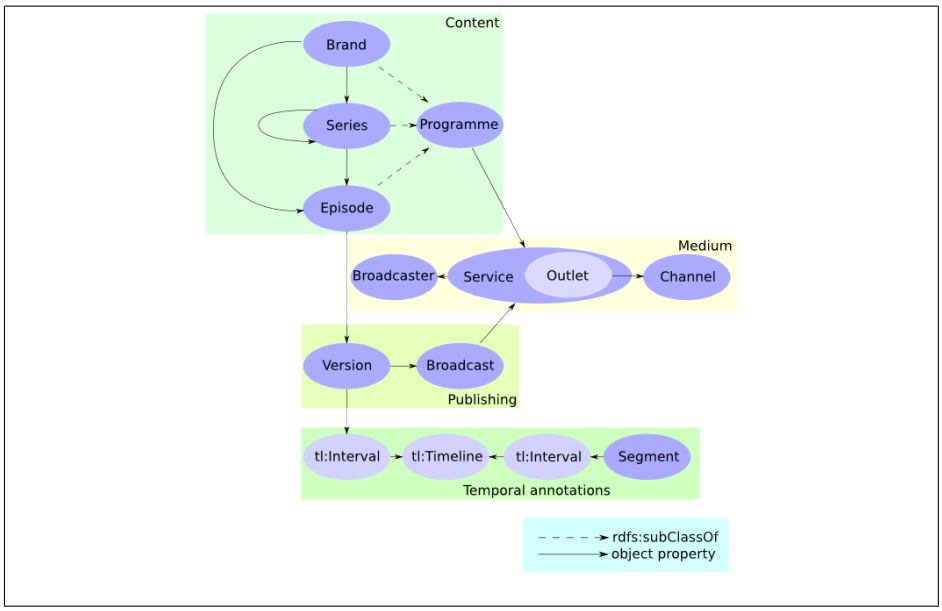

Figure 2: "BBC Programmes" ontology (http://bbc.in/peABAr) which underpins the BBCs information systems, and provides one of the key semantic building blocks of the BBC Stories ontology.

Despite some common design features and motivations, metadata standards such as MPEG-7 seem primarily intended to facilitate the transmission and playback of discrete units of multimedia data, whereas the formal characteristics of the BBC Stories and OntoMedia ontologies seem more influenced by the requirement to "capture and describe knowledge that is implicit in the context of the given media unit"(Jewell et al., 2005, p.1). For example, although MPEG-7 uses XML-based schemata internally to enable some support for both "low-level" semantic annotation within a single monolithic document, it does not support semantic links to external resources in a way that could enable at least conceptual, if not yet practical, compatibility with the distributed, layered approach to knowledge representation of the Semantic Web and RDF/XML (Stamou et al., 2006) (see figure 3).

Table 1 . Comparison of MPEG-7 and the Semantic Web.

Feature	MPEG-7	Semantic Web
Syntax	XML	XML/RDF
Schema/ontology language	MPEG-7 document description language (DDL)/XML Schema	RDF Schema/Web Ontology Language (OWL)
Composition	Monolithic	Many small layers
Extensibility	Aiming at completeness	Designed for extensibility
Multimedia ontologies	Part of the specification	Third party
Linking into media items	Part of the specification	Media dependent, incomplete
Available tools	None; not even a complete parser	Open-source tools
Real-life applications	None available	Mainly RDF with a few RDF Schema/OWL

Figure 3: A comparison of MPEG-7 and the Semantic Web as approaches to storing and exchanging multimedia metadata (Smeulders et al., 2000, p.1353).

Although the BBC Stories ontology represents the state of the art in expressing relatively abstract, "high level" conceptual knowledge about TV content in a machine-readable format, it is no less costly or complex to generate that knowledge in the first place. Noting advances in Computer Vision (CV)-based text, facial and feature recognition, inferences from audio processing, and myriad automated content analysis methods (Brunelli et al., 1996), researchers accept that a degree of human involvement is still necessary (Diakopoulos, 2009) in bridging the "semantic gap"[21]: the "lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation" (Smeulders et al., 2000, p.1353). The depth and breadth of the "semantic gap", and the necessary degree and quality of human involvement can be seen as proportional to the level of semantic richness required for the intended application (Davis, 2000), as well as dependent on what is understood by participants to to be "the given situation".

4.3.3 From Semantics to Pragmatics

Semantic Web researchers have emphasised the possibilities for enhancing content recommender systems by linking large semantically rich resources such as DBPedia with TV schedules and user data from Social Networking

[21] See appendix C.

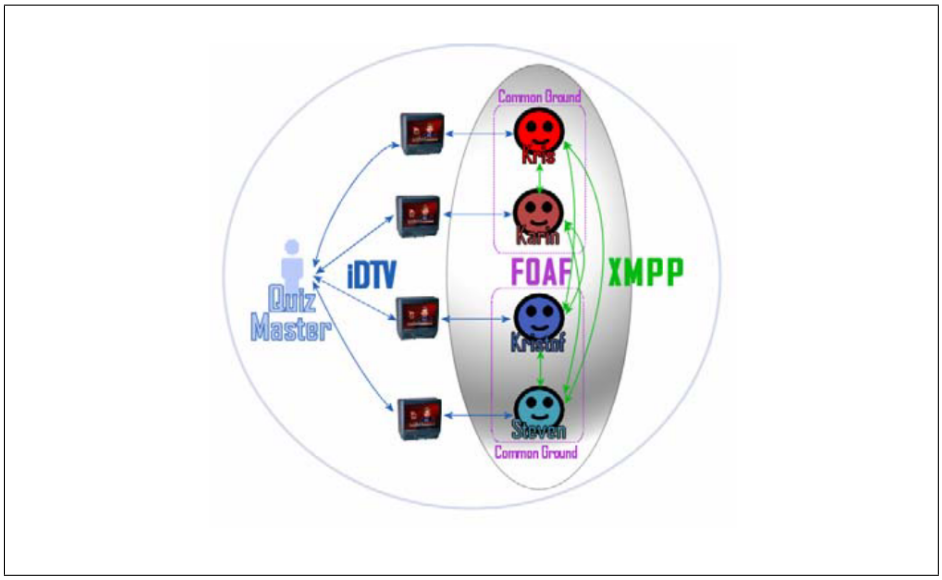

Figure 4: Telebuddies: "Using common ground to define manageable user groups" (Luyten et al., 2006, p.3).

sites (Aroyo et al., 2009), (Schopman et al., 2010), enabling innovation of new kinds of TV services. For example, Luyten et al. propose the "Telebuddies" system (see figure 4) in which viewer's "[s]hared characteristics are used to create a common ground between spectators." (Luyten et al., 2006). "Telebuddies" whose relationships are inferred from indexing and associating statements about their "name, address, hobbies, [and] hair color" found in their FOAF, profiles[22] are grouped and invited to join an "interactive" TV game show, where the predictable dramaturgy of the gameplay can be mapped onto an a-priori "interaction script".

In their description of the Telebuddies service (see figure 4), Luyten et al. use the term "common ground" to describe the knowledge the proposed Telebuddies system builds up about users by gathering ever-more detailed marketing information about their relationships and respective consumption patterns through logging highly constrained interactions such as "friending" or choosing to watch the same programme.

In a paper that looks at communication around content "in the wild" of online

[22]FOAF (Friend Of A Friend) is a Semantic Web ontology for describing people and relationships (Brickley & Miller, 2005).

interaction, Shamma et al. (2007) criticise approaches to media analysis that try to "close the semantic gap" between "low-level" and "high-level" representations of media objects, suggesting instead: "that media is best understood through the contexts in which it is used, and thus that research focus should shift in focus from semantics to pragmatics." (Shamma et al., 2007, p.276).

The pragmatic approach shown in this and related work (Liu et al., 2007), (Shamma, 2010) treats each screening of the video as an event that can be observed, rather than as a distinct media object that must be analysed. Mittell (2001) offers a cultural reading of television content, tracing the shift from TV as an event to be experienced to the TV show as a cultural object to be understood to the advent the Digital Video Recorder (DVR) enabling people to save TV shows and watch them later without the gaudy[23] intrusion of adverts[24].

4.3.4 The Technosocial Event

To provide a balance of pragmatics and semantics in gathering conversational data for comparison with structured metadata, several Social TV research projects and technologies provide useful precedents and guidelines.

Looking to live TV as the most event-like of televisual experiences, Shamma et al. (2009) studied large volumes of data culled from the "status update" service Twitter[25], in order to examine the relationship of tweets with relevant "hashtags"[26] to the programme structure and content of the 2008 USA presidential debates. Analysing the structural qualities of the data, rather than the elusive semantics, they were able to use onset detection methods (Bello et al.,

[23] Or years later, anachronistic and nostalgic.

[24] He also attributes the idea that TV only became regarded as an "artistic" medium recently to similar reasons, comparing the episodic, low production values of TV series until the late 1990's to the way the works of serialised authors such as Dickens and Tolstoy, were not regarded as "literary" until collected in bound editions years after their first publication as ephemera.

[25] http://twitter.com

[26] The Twitter platform allows uses to indicate loosely defined "topics" in their tweets using "hashtags", meaning a "#" symbol followed by a single word as part of their 140 characters of text. Users can then perform searches or "aggregations" of tweets with these hashtags. Another mechanism of addressing topics or people in Twitter is the use of the "@" symbol to denote the addressee of a tweet. For example, a tweet with "@saul" in the text would be seen as "addressed to" the user "saul", prompting an "alert" to that user and potentially opening up a dialogue via twitter (Honey & Herring, 2009). In this case Shamma et al. (2009) collected tweets marked with "hashtags" such as #obama, or #mccain, or those "directed" @obama.

2005) to infer highly accurate debate topic boundaries from frequency of tweets evident between debating points, and compare them to the structured closed caption data provided by CSPAN[27]. However, the CSPAN data showed only the most broad, a-priori topic boundaries. Compared to the rich interactional data recorded between co-present viewers in more intimate studies (Oehlberg et al., 2006) (Harboe et al., 2008b), the Twitter data seemed lacking in observable contextual detail required to enable a detailed analysis or even ascertain the shared communicative groundings beyond those inferred through hashtag use. For example, during the presidential debates, some of the most frequent tweets appeared as unintelligible streams of numbers, which, because of the tendency that the same tweets were marked with the hashtag "#drinking", the researchers interpreted to be scoring in some kind of drinking game (Shamma et al., 2009), but without further contextual markers, the researchers were unable to speculate further on the detail, rules, nuances or the social dynamic of the supposed game.

Some Social TV systems have been specifically developed to capture detailed, explicit annotations from TV viewers via a "Watch and Comment" paradigm (Cattelan et al., 2008) . However, most of these tools and approaches seem focused on semantics, and appear complicated and oriented strongly towards the task of annotating content rather than watching TV during a social event[28].

[27]The Cable-Satellite Public Affairs Network (http:c-span.org) shows all congressional and federal official broadcasts on all cable and satellite networks in the USA.

[28]See appendix C

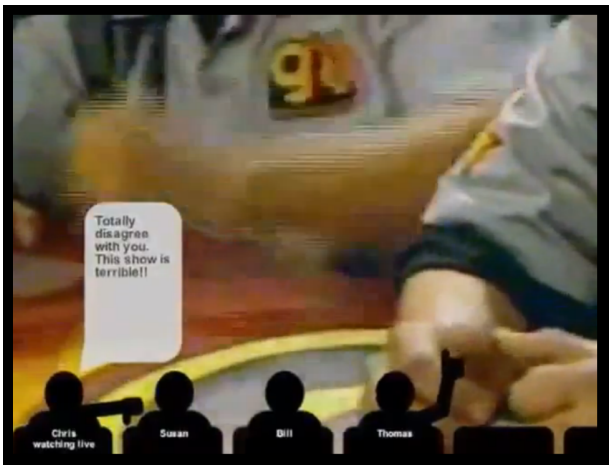

Figure 5: A screenshot from the user interface of the CollaboraTV system (Nathan et al., 2008).

In contrast, Nathan et al. (2008) developed the CollaboraTV system to enhance and stimulate the interaction of TV remote viewers as if they were co-present. CollaboraTV enabled viewers to comment asynchronously and iteratively alongside a timed video track via an intuitive “virtual audience” interface, and see others’ comments pop up at time-coded moments in the video, intended to appear as if viewers were watching together (see figure 5).

Of the Readily available Social TV annotation systems[29] most seemed geared to supported groups of remote viewers, with combined annotation and viewing interfaces on a single screen, rather than providing a space for co-present orientation to a shared display of content. However, some proposals for ”second screen” interfaces (Cesar et al., 2008a), (Kramskoy, 2011), suggested methods of separating the content presentation interface from the annotation interface in ways that could enable multiple co-present viewers to co-annotate video using the shared focus of a single display screen.

Generic chat tools and “backchannel” systems[30] for event annotation also

[29] Very few of the Social TV systems reviewed in this paper were available to re-use and modify under Open Source licenses.

[30] A “backchannel” has a variety of meanings in non-verbal communication and information science. In this instance it is used to refer to the tendency, particularly during conferences in the technology community, to use chat tools or Twitter to communicate between an audience at a live

provided a “main screen” for shared, co-orientation (Weisz et al., 2007) to event-like content or situations, however most seemed specific to one simple function[31], or had complex display and interaction mechanisms[32].

To construct a specific situation, likely to elicit fluent, mediated, co-present interaction in response to a shared content display, “The Heckle Tool” was developed and tested to gather preliminary interactional content-related data for later comparison with structured, broadcaster supplied metadata for the same media content.

5 Materials

Two sources of data were required for this study: data from conversational interaction between viewers of a video, and structured, broadcaster-supplied metadata for the same video.

5.1 The Heckle Tool

5.1.1 Design Process

The Heckle tool was developed from an existing prototype for a live event video annotation system by media art collective *The People Speak*[33], who permitted re-use and modification of their Heckle system[34] for the annotation of TV media as part of this research. Throughout the development and user testing process, requirements for a second iteration were gathered[35].

event, often providing a critical or auxiliary commentary to the current speaker (McParland, 2002)

[31] Such as http://www.backnoise.com or http://www.todaysmeet.com, text-only web-based chat systems.

[32] Such as http://www.wiffiti.com which enables multi-modal event annotation, but either requires logging in via Social Networking and Social media services, or has a complex and inflexible image upload interaction and visualisation process.

[33] http://http://thepeoplespeak.org.uk

[34] See appendix E.

[35] See appendix **??**.

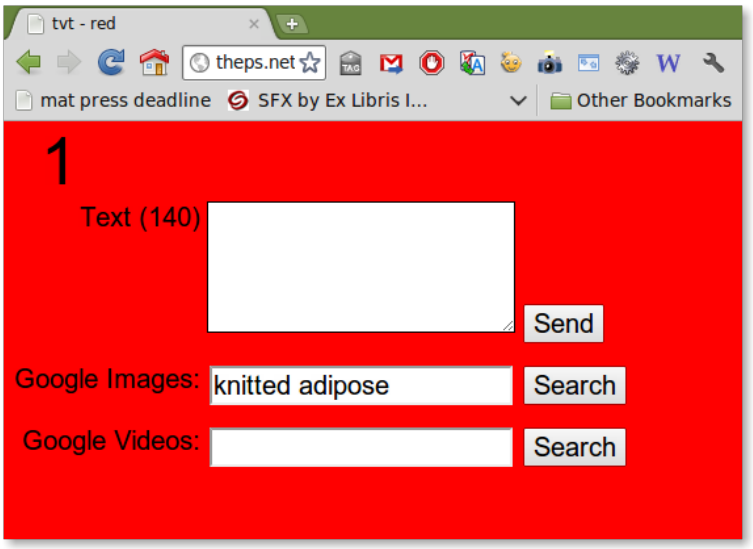

Figure 6: Heckle client annotation interface for the user who chooses red to identify themselves on the shared screen.

5.1.2 System Architecture

The first Heckle system prototype was built on a client-server architecture, on a Linux[36], Apache[37], MySQL[38], PHP[39] (LAMP) software stack. The Heckle system integrates various third party web service APIs to provide image and video annotation functionality[40].

5.1.3 Annotation Interface

The Heckle system provides a simple web interface for participants who are using laptops, tablets or any device with a web browser to type in 140 characters of text or perform a quick Google search for images or youtube videos, before pressing “send” to show them on the “display screen” (See figures 6 and 7). As the interface loads for the first time, users are asked to choose a colour to identify their “heckles” when they are sent and displayed.

[36] http://ubuntulinux.org
[37] http://apache.org
[38] http://mysql.org
[39] http://php.net
[40] In this trial the Google Search API was used to provide interface to Google Image Search: http://api.google.com.

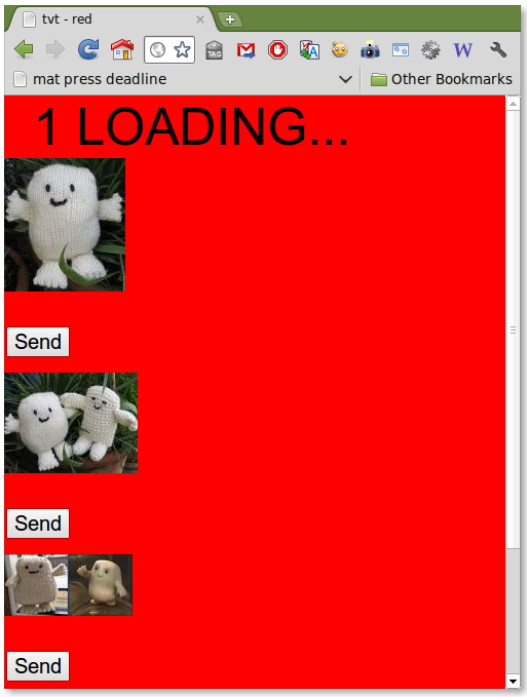

Figure 7: Heckle client annotation interface showing search results for “knitted adipose”.

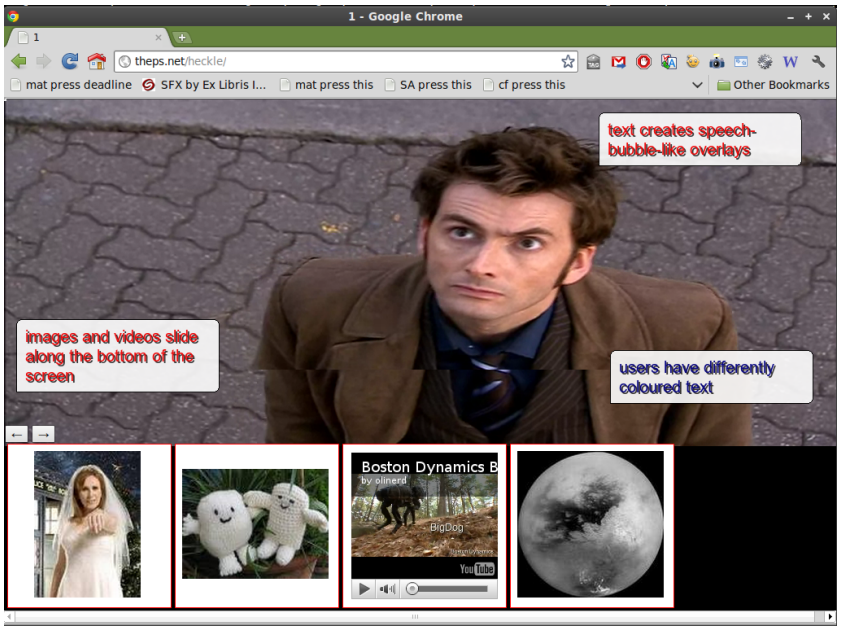

Figure 8: Heckle system display screen showing an episode of Dr Who running in the background with demo annotations.

5.1.4 Display Interface

The display screen shows full-screen video (in this case, an episode of TV show *Doctor Who*), alongside texts, images and video clips sent by viewers in a full-screen browser window. The images and videos sent by users have a coloured outline, and text-bubbles are coloured to indicate which user posted them (see figure 8).

Images and video clips run underneath the video in a “media bar”, while text bubbles posted by users drop onto the screen in random positions, but can be re-arranged on the screen or deleted by a “facilitator”.

5.1.5 Facilitator Interface

The facilitator interface enables a facilitator to click and drag text bubbles around the display screen arbitrarily. They are also able to clear the media bar, the text bubbles, or the entire screen with a single click (see figure 9).

This enables the facilitator to reconfigure the detail of the flow of information

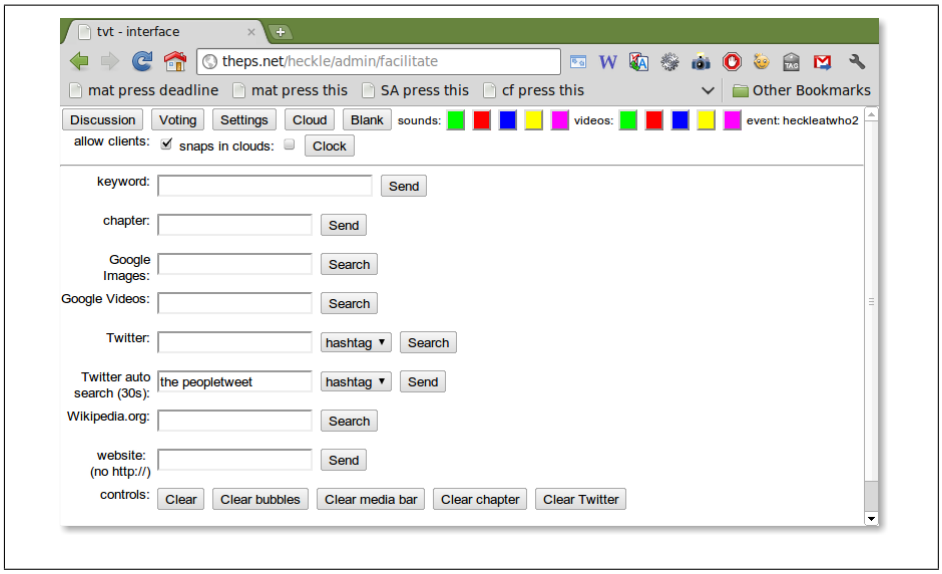

Figure 9: The Heckle system's administration/facilitation interface.

on the screen, in response to live requests and feedback from the viewers[41].

5.1.6 Data Collection and Review

All "heckles", images, video and text are recorded in a database along with an absolute time-stamp and colour information to indicate which user posted the heckle, and at which point. The absolute time-stamp, allows each heckle to be matched with a specific time-code in the video.

A "data review" visualisation enables the display of all heckles so far in a linear, colour-coded timeline for analysis and review (see figure 10).

[41] This method of using a human operator to perform functions that, in production software, would be automated is sometimes known as "Wizard of Oz testing" (Rice & Alm, 2007), (Cesar et al., 2008c), and has been shown to be effective in early stages of Social TV (and other) software development to maximise the effectiveness of user tests, without requiring a huge investment of development time in complex information presentation functionality that may or may not achieve the intended result without the initial verification of user feedback.

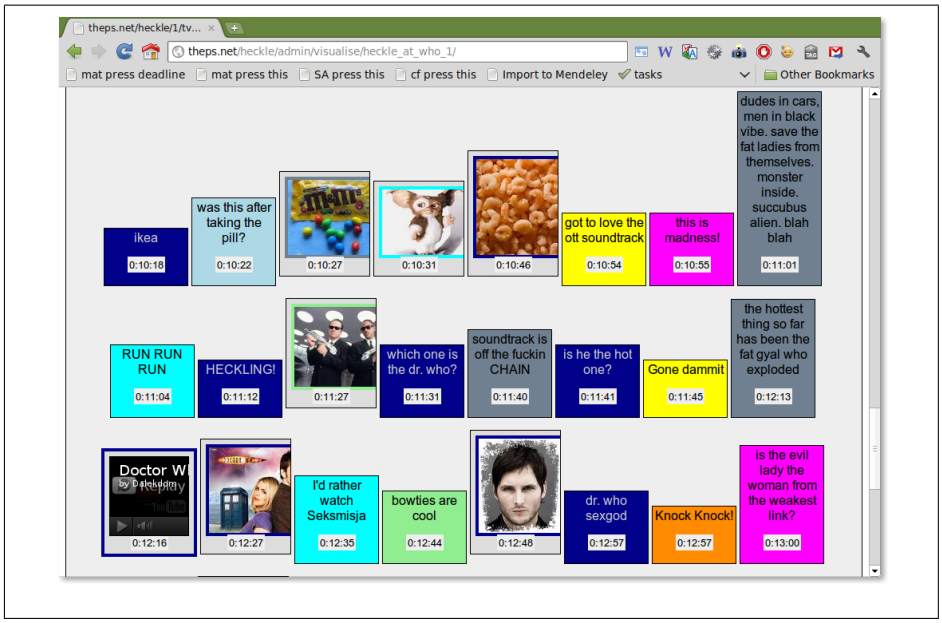

Figure 10: The Heckle system's data review mode showing "heckles" from users at the second test.

5.2 Broadcaster-supplied metadata

5.2.1 BBC Stories annotations

The BBC Stories project's (Harris, 2010) highly detailed semantic annotation of Season 4, Episode 1 of *Doctor Who*. "Partners in Crime" (Strong, 2008) provided a benchmark of highly structured data against which to evaluate the relative content-relatedness and structure of TV viewers' conversations.

The annotation data were provided as an easily readable and editable N3 syntax formatted (Berners-Lee, 2006) RDF file built on the BBC Stories ontology[42]. As shown in figure 12, the metadata is composed as a series of statements about the ontological relationships of subjects, predicates and objects, or "triples" relating to the narrative, character or production elements in the show. These triples were loaded into an instance of the 4Store RDF database[43] to enable easy semantic queries that traverse the data's graph-like structure.

[42] http://www.contextus.net/stories/
[43] http://4store.org/

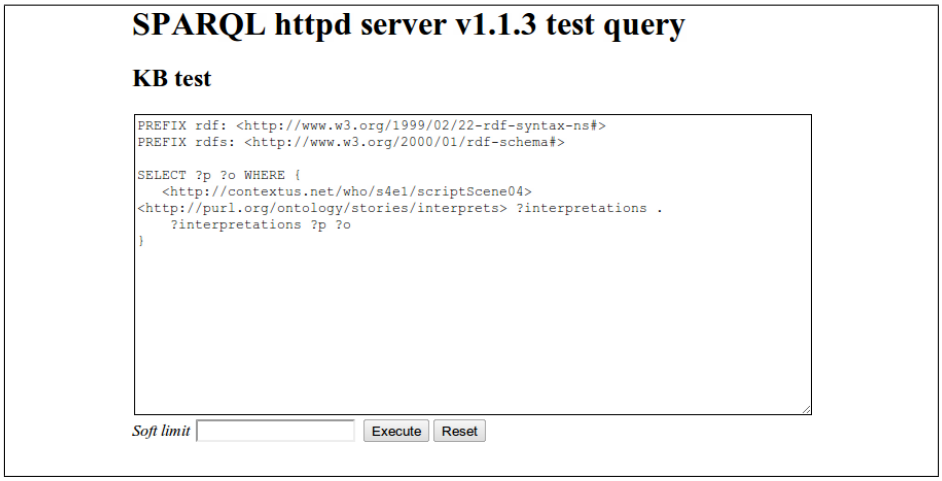

Figure 11: A query on 4Store's http SPARQL endpoint, used to generate a representation of scene 4 of the episode.

```
:scriptScene03 a stories:Interpretation;
        rdfs:label "Script - Scene 3";
        rdfs:comment "THE DOCTOR steps out of the TARDIS. Sets off. On a mission.";
        rdfs:resource <path/to/encodedscript/#sceneX>;
        stories:interprets :building_entry;
        stories:retracts :doctor_in_tardis;
        stories:source :script.

:scriptScene04 a stories:Interpretation;
        rdfs:label "Script - Scene 4";
        rdfs:comment "DONNA walking along, left to right, through COMMUTERS.";
        stories:interprets :building_entry;
        stories:source :script.

:scriptScene05 a stories:Interpretation;
        rdfs:label "Script - Scene 5";
        rdfs:comment "THE DOCTOR walks along, right to left, through COMMUTERS.";
        stories:interprets :building_entry;
        stories:source :script.
```

Figure 12: An extract of the n3 RDF data loaded into 4Store relating to scene 4 of the script.

5.2.2 BBC Stories object data collection

Using 4store's http SPARQL endpoint to ask some simple semantic queries (see figure 11), an XML representation of all the "objects" in the BBC Stories annotation was generated that related to the "subject" of each scripted scene of the episode via the predicate "interprets" (see figure 13).

This list of objects, subdivided into sequences of time-bound scenes within the video provided a sufficiently detailed dataset for a series of simple, pragmatic analyses.

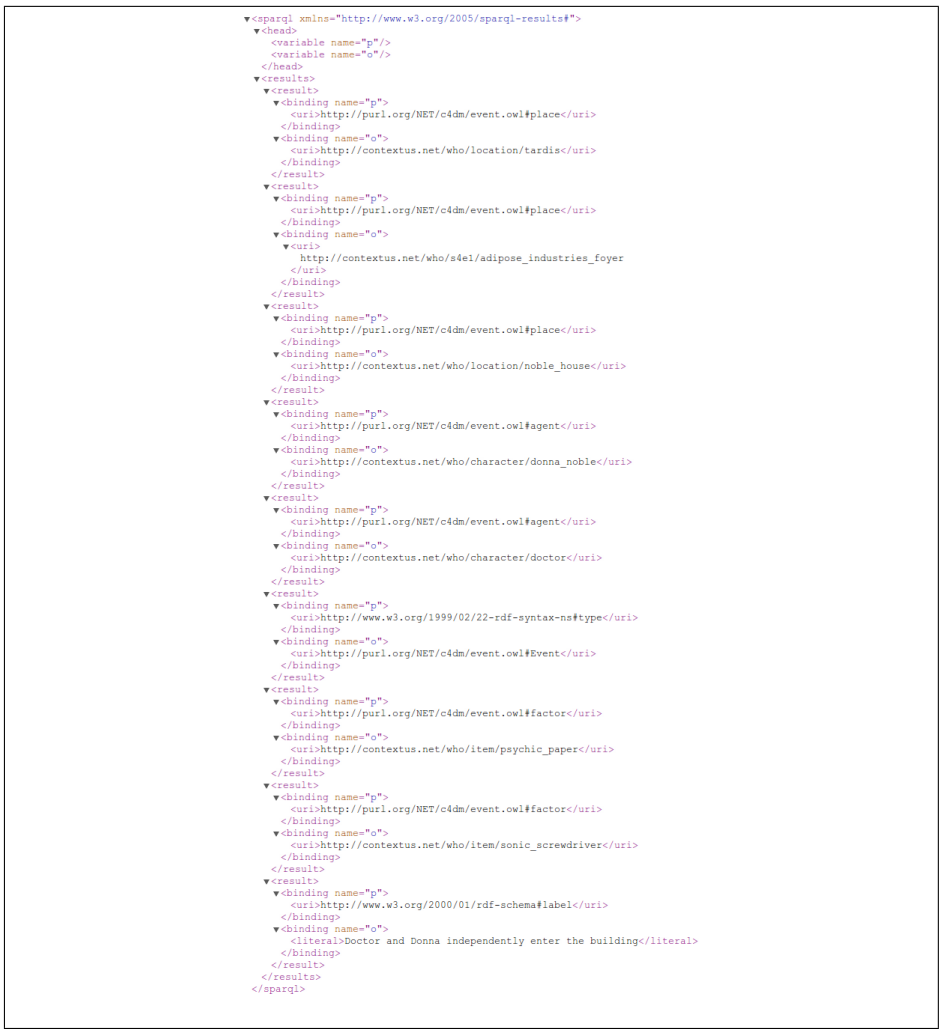

Figure 13: The results from the query in figure 11 showing all the objects related to scene 4 of the script by the predicate “interprets”.

6 Methods

6.1 Participants

Two groups of eight volunteers were selected from respondents to invitations to participate in the trial. Each group consisted of colleagues from the same workplace with existing social connections.

6.2 Site and Equipment

The trial took place in an informal screening environment, intended to emulate the context of "the event" mode of co-present television viewing: friends assembling at home to watch a specific sports programme or film together. A comfortable sofa and bean-bags were set up in a broad semi-circle facing a large projection screen, with drinks and snacks set up on a bar to the side of the seating area.

All participants were asked to bring their own laptops, web-enabled tablets or mobile phones. Power sockets were made available behind each seat, and a wifi network was provided for fast Internet connectivity.

Behind the sofa arrangement, a "facilitator" was stationed with a laptop to control the audiovisual presentation and the Heckle software system. The facilitator's laptop had a dual-screen monitor configuration, showing the Heckle system's display interface on the video projector, and the administrative interface on the built-in laptop screen. The facilitator was close enough to the sofa to hear and be heard by any member of the viewing group, and to be able to respond to spoken or "heckled" requests by viewers to, for example, use the administration interface to re-organise the flow of information on the display screen.

6.3 Procedure

Before the test, the facilitator checked that people were able to use the network and asked participants to visit the web address of the annotation interface, and choose a colour. They were then informed that once the video started, they should feel free to get up, socialise with each other, move around and help themselves to food and drink at any point, and to talk if the heckle interface was too cumbersome.

With the verbal consent of participants, an unobtrusive miniature video camera was set up on a tripod at the back of the room to capture the overall scene, vocal exchanges and people's other interactions and behaviours during the test.

The facilitator was equipped with a text editor on the same laptop being used for the Heckle system, to take notes on the state and functioning of the system, and to record observations about people's interactions and behaviours, and informal feedback from participants during and after the test.

Once the video began, all texts, images and video clips the participants sent to the display screen were recorded in the Heckle system's database.

The first test aborted after fifteen minutes due to technical failure with the Heckle system, however, sufficient interaction data were successfully collected to enable analysis. The problem was solved for the second test, but only fifteen minutes of data from each test is presented here to enable comparison between the two sessions.

6.4 Analysis

6.4.1 Data transformations

The heckle data log timestamps were shifted backwards several minutes to compensate for false starts to the test, and several seconds to compensate for network latency issues having caused slight delays between participants posting and heckles being received and displayed.

Duplicate heckles, which had been unintentionally sent when participants reloaded their web clients[44] were removed from the data.

6.4.2 Lexical alignment

Nouns, and noun phrases in the heckle data were counted and extracted into lists for each test. An equivalent list of objects were extracted from the BBC Stories metadata by removing XML markup from the results of the SPARQL query, leaving only nouns referring to annotated objects.

To gain a first insight into overlaps in the vocabularies of their domains of reference, a simple frequency analysis was performed on both lists to find overlaps in frequently mentioned words.

6.4.3 Contextual ground

To find evidence of how dependent interactions are on either the context or on the content of the video, anaphora in the heckle data were counted. Along with previously counted nouns and phrases, each heckle was typed into two broad categories: references to events and characters in the video of Dr Who, and references to the context of the test, such as other participants, the organising of activities in the space, and the heckle software itself.

Ambiguous or phatic expressions were not typed in either way. Phrases which referred to multiple objects, or that referred to objects relating to both the video and the present context were counted and typed for each reference. Images and videos that seemed to reference either the video or the context clearly were also counted.

Heckles relating to prior heckles were typed as relating to the video if the heckle to which they replied was clearly video-related, but only with one level of depth. Unless this referring-heckle also explicitly referred to the video, any further references relating to it would be typed as context-related[45].

[44] This flaw in the Heckle software is solved in the subsequent version.

[45] This is debatably equivalent to the method by which objects were selected from the RDF data described above, which equates to a single traversal of the graph, via the predicate “interprets”

6.4.4 Annotation density

To gain a measure of the comparative density of references per scene or event, the video was reviewed to find time-codes for the beginning and end of each scene as represented in the BBC Stories metadata.

Scenes that were present in the script but had been omitted from shooting were removed from the data set. Scenes that were more than a few seconds long, but had been moved or inter-cut with other scenes during editing and production process were re-aligned with the order in which viewers had watched them when heckling during the tests.

Each heckle was then assigned scene markers according to rough correspondence of heckle timestamps to time-coded scene segments, providing a list of heckles per scene. Cross-referencing this list provided a count of nouns, anaphora and noun phrases per scene.

The BBC Stories metadata was queried to produce a count of objects referenced per scene with which to compare the scene-aligned heckle data.

6.4.5 Conversational markers

A simple visualisation of the heckle stream was built as an add-on to the Heckle system, enabling a visual overview of the recorded conversation flow along with time codes for each text, image, or video heckle (see figure 14).

Although no formal Conversation Analysis was applied to the heckle data, techniques derived from CA were used to find conversational markers, that could indicate the relevance and feasibility of performing a Conversation Analysis on heckle data in the future.

These techniques were employed to look for basic evidence of conversational behaviour in the form of turn-taking and sequences of heckling between participants. Within sequences, more complex conversational tropes such as self-repair, other-repair and repair initiation were also explored. Sequences found in the flow of heckles were first typed in terms of their contextual refer-

Figure 14: An example of the Heckle visualisation built to provide a visual overview of the conversation during a screening and for later analysis.

ence to either the video or to the interactional context, then examples of topic tying were sought in order to analyse the detail of how and if contextual relevance of topics was negotiated through heckling.

7 Results

7.1 Lexical alignment

Word frequency analysis of nouns and noun phrases in each data set shows a degree of overlap in the domains of reference, with “dr who” or “doctor”, and “adipose” and “fat” being the amongst the most relevant terms in each data set (see Table1).

Although the crude method of this analysis breaks apart phrases such as “sonic” and “screwdriver” into separate terms, conceptual overlaps such as “fat” and “adipose” are still apparent, as is the consistent lack of overlap in common terms of reference for what are seemingly the same objects.

For example, in the BBC Stories metadata, “foster” (the evil nemesis in this episode) and “lady”, in the heckle data are at equivalent frequency levels. In the heckle data this character is referred to twice as “the anti-fat lady”, and once as “the evil lady”. Similarly, even though she is a central character in a popular TV series, who says her own name multiple times in the script, the name Donna Noble which appears just under “doctor” in the BBC Stories data is referred to twice in the heckle data by the actor’s name: “tate”[46].

Only the Doctor, the central character, is referred to by the same name in each data set, although even there, the names are spelled and abbreviated differently.

The heckle data seems to contain objects which are relevant to organising local interaction, such as the names of one participant “zoba”, and references to a “fag”, from a heckled conversation about organising a collective cigarette

[46]The actor Catherine Tate plays the character Donna Noble in the video.

BBC Stories metadata		Heckle data	
Term	Occurrences	Term	Occurrences
adipose	55	who	18
doctor	51	fat	10
donna	49	dr	10
noble	47	pills	6
industries	37	screen	5
house	25	pill	4
psychic	21	god	4
paper	21	zoba	3
foster	21	tom	3
sonic	15	one	3
screwdriver	15	lady	3
squad	14	fag	3
collection	14	diet	3
tardis	12	alien	3
staceys	10	weight	2
roger	10	tate	2
journalists	10	soundtrack	2
independently	10	singer	2
gizmo	10	seksmisja	2
foyer	10	safety	2

Table 1: Simple word frequency of top 20 words in each data set.

break. From the heckle data, only 5 of the 20 most frequent words[47] seem to relate more clearly to the context than the video, the rest are more easily identifiable as related to themes from this episode of Doctor Who.

[47]The words:“screen”, “god”, “zoba”, “fag”, and “seksmisja” in the data can be attributed to conversations and interactions that relate to contextual markers.

7.2 Contextual ground

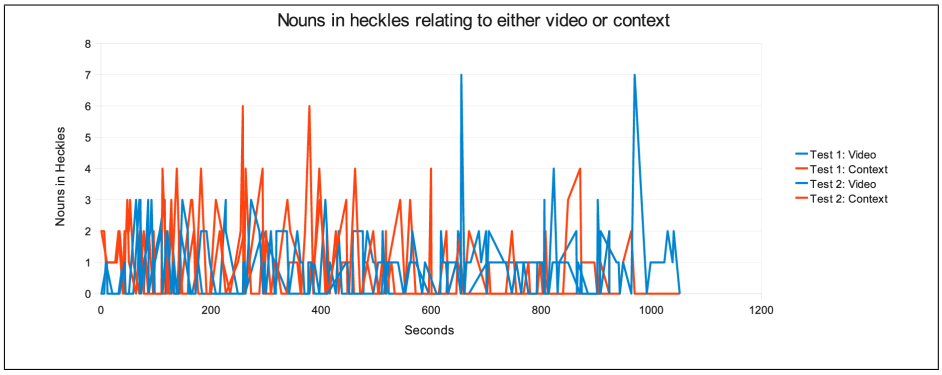

Figure 15: Graph of nouns, noun phrases and anaphora from the data of two Heckle tests referring to objects in the video, or to to objects the interactional context

In the opening 15 minutes of the first test, heckles were divided equally between those grounded in the video and those referencing context. In the second test, twice as many heckles seemed to be grounded in the video as opposed to the context, with 20% of those referring to both grounds.

The patterns of reference to these two grounds is explored by plotting the nouns, anaphora and noun phrases from the two tests, divided into references to what is happening in the on-screen video, and the context of the situation at hand (see figure 15).

The frequency of references to video and context from both test groups seems roughly balanced, with some suggestion that in both tests, intense clusters of heckling activity correlate at certain points during the screening, but not clearly in relation to one or the other ground, although the peaks of the video-related heckling seem slightly more frequent and sharper.

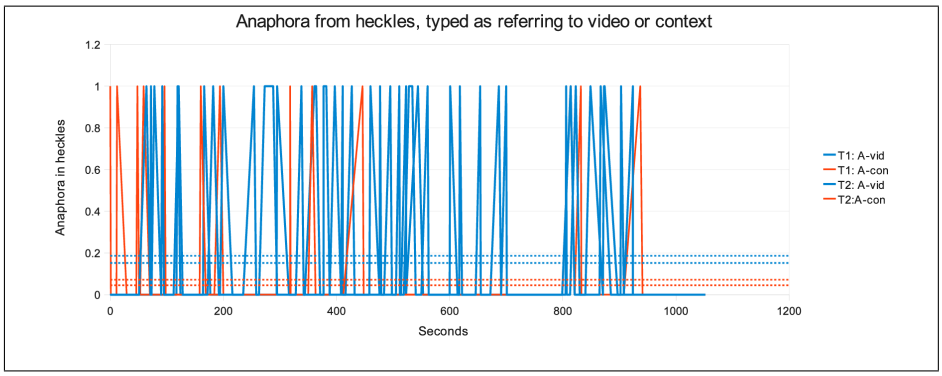

Figure 16: Graph of anaphora in heckles relating to prior turns throughout two Heckle test sessions relating to objects in the video, or to objects in the interactional context.

In both tests, 23% of object references in all heckles are themselves references to prior heckles. References to objects in prior heckles appear to be equally related to either the video or the context.

However, as shown in figure16, although only 16% of all heckles contain anaphora, which constitute relatively strong contextual markers for shared viewer orientation, 77% of nouns referenced by anaphora in all heckles are typed as relating to the video. Also it seems that as the screening progresses, contextually grounded anaphora give way to a higher frequency of video-related heckles.

7.3 Annotation density

The BBC Stories metadata for the first 15 minutes (or 47 scenes) of the show included 335 references to objects. For the same time period, each Heckle session yielded on average 207 references to nouns, anaphora and noun phrases.

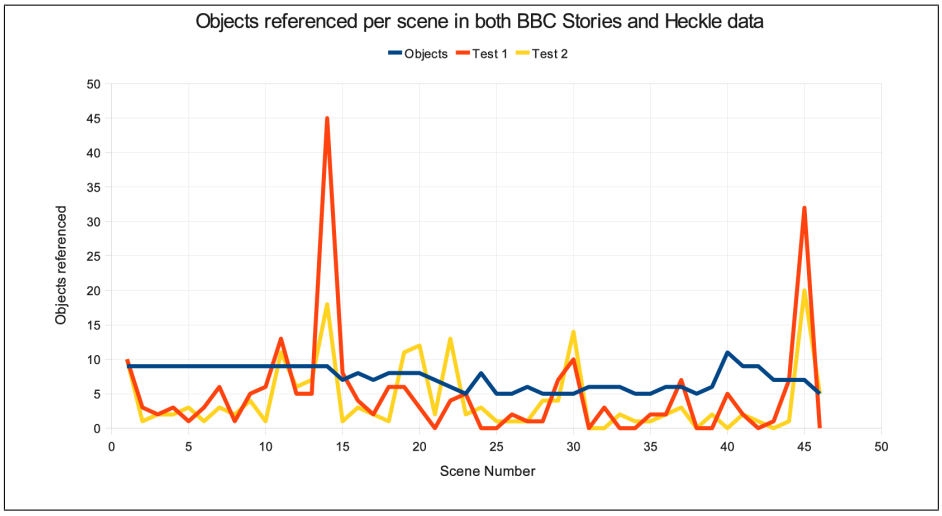

Figure 17: Graph of objects referenced per scene in both heckle data and BBC Stories metadata.

Plotting the number of annotations per-scene in both sets of data (see figure17) showed very different frequencies of reference. Whereas the BBC Stories metadata shows a relatively even number of annotations in each scene, some of which are only seconds long, the Heckle data shows a far more varied distribution of references in peaks and troughs.

7.4 Conversational markers

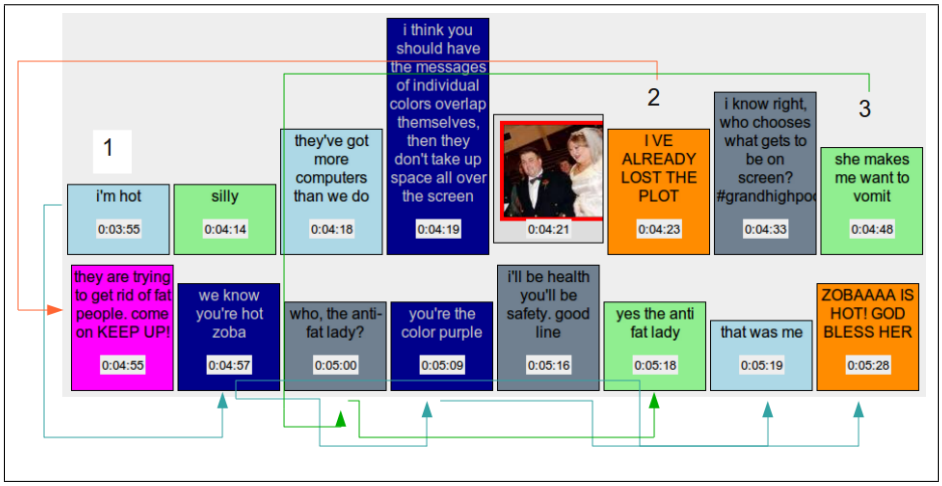

Figure 18: A set of interleaved sequences from Heckle test 1.

Turn taking sequences and repair structures were evident throughout the data. For example, figure18 shows three sequences, running in two rows from left to right. The third sequence (numbered 3) is initiated at 04:48 by the green user using an ambiguous anaphora “she makes me want to vomit”. At 05:00, the grey user marks the ambiguity, other-initiates, and offers a candidate for the self-repair: “who, the anti-fat lady?”. The Green user completes the sequence by picking up the preferred explanation: “yes the anti fat lady”.

A more complex sequence (numbered 1) is initiated at 03:55 by the light blue user heckling “I’m hot”, as a “status-update” type of heckle. This is picked up in the following line by the dark blue user as an intentional misunderstanding: “we know you’re hot zoba”. The same user continues the sequence, identifying zoba as the purple user “you’re the colour purple”, not the light blue user as presumed. The sequence is then picked up by the light blue user again, correcting the misattribution: “that was me”. Several turns later, the purple user picks up the sequence again, confirming her identification: “I am purple” (see figure 21). Subsequent exchanges between the dark blue and purple user indicate that they make use of the function of this sequence: having identified one another by colour.

Another sequence (numbered 2) is initiated at 04:23 “IVE ALREADY LOST THE PLOT”. This is picked up by the purple user at 04:55 “they are trying to get rid of fat people, come on KEEP UP!”. The mirroring of all-caps syntax suggests that these are sequential, seemingly confirmed by the reply from the orange user, which also forms part of sequence 1, at 05:28 “ZOBAAAA IS HOT! GOD BLESS HER”. This reply simultaneously picks up on the outcome of the first sequence in which zoba is called hot, and supplies a response to the purple user’s offered explanation of the plot.

The sequence in figure 18 also shows some evidence of conversational structures such as self-initiated self-repair “you’re the colour purple” (noting a mistake in the blue user’s own previous turn), and, in the same phrase heckling about the misidentification of zoba provokes a self-initiated other-repair from

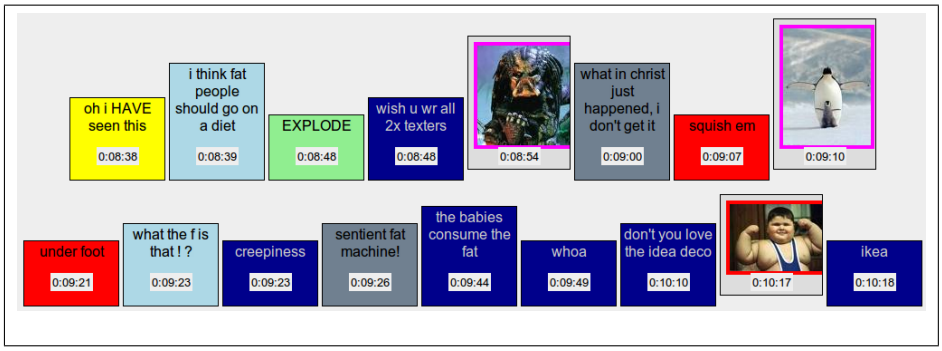

Figure 19: An example of self-initiated self-repair in a sequence of heckles.

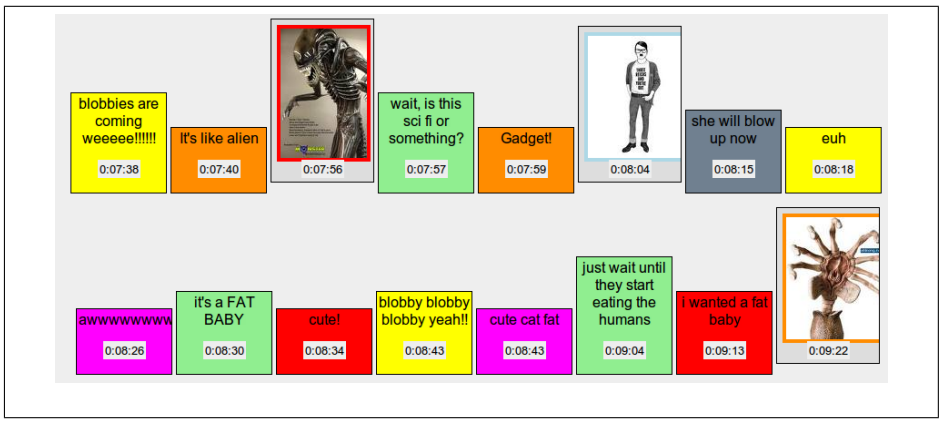

Figure 20: An example of other-initiated self-repair using images from the film Alien

the light blue user at 05:19: “that was me”.

A more clear-cut example of self-initiated self-repair is evident in figure 19, in which the dark blue user mistypes “don’t you just love the idea deco”, and then self-repairs “idea” to “ikea”.

Something like other-initiated self-repair also seems to take place in multi-modal communications. In figure 20, the orange user at 07:40 says “it’s like alien”. At 07:56, the red user offers an illustration, or a tacit question “you mean like this?” by heckling an image of a prop from the film Aliens. At 09:22, the orange user seems to reply with an image of a different variety of alien from the same film, that more closely resembles the “Adipose” aliens depicted in the video.

Figure 21 shows a continuation of the heckle stream from figure 18, in which several topics are negotiated and hybrid topics emerge that mix contextual cues

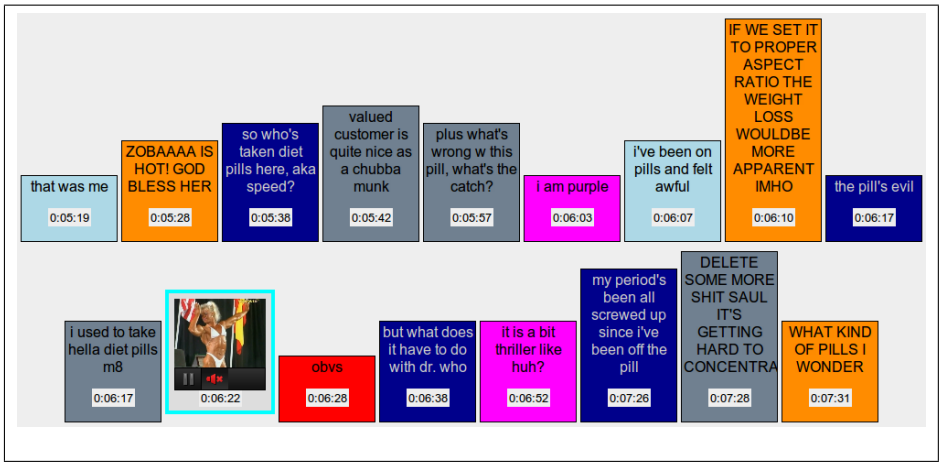

Figure 21: A continuation of the heckle stream from figure 18, a series of interlinked turns, tying exchanges into hybrid topics

and references to the video of Dr Who.

The dark blue user initiates a new topic, only partially related to an element of the video narrative at 05:38, linking it to future utterances with the tying structure “so”: “so who’s taken diet pills here, aka speed?”. At 05:57, the grey user initiates a parallel topic, with an anaphora relating the topic back to the video more closely: “plus what’s wrong w this pill, what’s the catch?”. In this case, “this pill” is the pill shown in the video. At 06:07, the light blue user responds in sequence to the first question (using the plural “pills”) “i’ve been on pills and felt awful”. Using the singular, tying to the parallel topic started by the grey viewer’s question at 05:57, the dark blue user then responds to the question about the video-related pill topic using the singular “pill” at 06:17 “the pill’s evil”. Simultaneously, the grey user responds to the initial question about diet pills: “i used to take hella diet pills m8”. A loud video of a dancing female bodybuilder is heckled onto the screen by the cyan user and after a short exchange discussing it’s relevance until 06:36, the blue user suggests a further related topic: “my period’s been screwed up since I’ve been off the pill”. At 07:31 The orange user then heckles “WHAT KIND OF PILLS I WONDER”, using the plural “pills”, seemingly tying the question to the light blue user’s earlier lack of clarity about the pills that made them feel awful at 06:07.

Although no formal CA methodology was applied in the analysis of the heckle data, structures of turn-taking, sequences, and various forms of repair were observed and evidence was found demonstrating the use of these structures to tie heterogeneous elements into topics for negotiation and use in conversation.

8 Conclusion

8.1 Discussion

8.1.1 Terms of reference

The results of a simple word frequency analysis have shown clear conceptual, if not lexical correspondence between the terms used to describe content in both formal broadcaster ontologies, and the practical ontologies of media content used in communication.

A similar observation was noted by Shamma et al. performing a term frequency analysis of a far larger corpus of Twitter data (Shamma et al., 2009), and the possibility of using thesauri to cross-reference vocabularies of content in practical communication with formal topic descriptions was suggested.

However, detailed observations on the microcosmic scale of two 15 minute conversations suggest that that these vocabularies of practice are very well adapted to match the imperative to understand and be understood in the situation at hand. For example, in the heckle data, the character “Donna Noble” is referred to as “tate”, or “the readhead”, and “Miss Foster” becomes “the evil lady” or “the anti-fat lady”. This lexical mobility in response to the need to communicate seems to mitigate against the likelihood of tractable stability and conformance in terms of reference between conversational and formal ontologies.

This simple analysis also demonstrated that conversations around media differ from formal media content ontologies in that they contain multiple refer-

ences to the context at hand, as well as to the themes and objects referenced in the content. However, it could be argued that had they been retained, the XML markup in the raw BBC Stories data could have been seen as references to the "context at hand", somewhat equivalent to context-related references evident in the heckle data.

8.1.2 Grounding in content

While seeking to evaluate whether synchronous or asynchronous video viewing can provide a "common ground" (Clark & Brennan, 1991) for communication, Weisz et al. coded text chat logs from a laboratory study in remote co-viewing of videos and claimed that 65% of the conversation was about the TV content or evaluations of it (Weisz et al., 2007). Harboe et al. coded audio chat between remote co-viewers of a Social TV system, and claimed to show that their phatic communications were effective because of the "shared context and common ground provided by the TV content" (Harboe et al., 2007), and even more so between groups that knew each other (having other shared common grounds). Weisz and Kiesler claimed that synchronous and non-synchronous co-watchers of video used text chat an equal amount, but that in their coded chat logs, asynchronous co-viewers "compensated" by talking about different grounds: their personal lives, and updates about what they were each watching (Weisz & Kiesler, 2008). Nathan et al. found that 75% of Likert-scale respondents strongly agreed that TV content provided a "common ground for conversations with friends and family". Within that sample, 37.5% used TV as a "conversation-starter", 37.5% to illustrate points in conversation with others and 18.8% to clarify values and opinions to others (Nathan et al., 2008).

The task of demonstrating the grounding of communication through co-viewing of TV content may have been orthogonal to the research aims of these Social TV projects. Furthermore, the research methods and technologies employed in these studies often seem inconsistent and, especially in the latter case, incompatible with a commitment to observation of the specific technoso-

cial situation. However, the otherwise tacit consensus in much of the Social TV literature, that TV content can be seen to provide a common ground for communication (Schatz et al., 2007), (Harboe et al., 2008b), (Aroyo et al., 2009), (Cesar & Geerts, 2011), is consistent with the findings here.

Measuring the relatedness of nouns, anaphora and noun phrases in the conversational data to either context or the media content indicated that there was a more or less equal balance between viewers orientation to each. However, the relatedness of 77% of anaphora in the data to the video content, and the relative prevalence of content-related heckles as the screening progressed indicated that the content became increasingly central to the social interaction, frequently requiring little more explanation other than “him”, “her” or “that” to be understood as the shared focus of attention.

8.1.3 Semantics from pragmatics

Shamma et al. (2009) assert that in content annotation, “semantics would be tied to the pragmatics of the annotating application” (Shamma et al., 2009, p.3). This is supported by the observed differences in the frequency of annotations per scene between the formal content metadata and the heckle data. In figure 17, the sharp peaks and troughs of the conversational annotation contrast with the relatively stable numbers of annotations per scene in the BBC Stories metadata.

This could be attributed to both the pragmatics of the tools: a text editor, in the case of the BBC stories metadata (Harris, 2010, p.24), and the Heckle tool in the case of the conversational data. However, The emphasis on the *tool* determining the semantics of the annotation seems to underplay the structuring contingencies of the communicative context. A detailed overview of the conversational qualities of the heckle data revealed very different imperatives than those of a solitary researcher sitting in front of a computer with a TV script.

8.1.4 Semantic drift

Structural evidence for the conversational nature of the heckle data was clear throughout: turn-taking, adjacency sequences, and even multiple forms of repair suggested that this, or similar data derived from similarly constructed situations would be amenable to Conversation Analysis techniques and related methods.

In the absence of a tractable metric for "sociability", this suitability of the data for conversational analysis suggests that "watching together" and heckling constitutes a "communication space" that can be read as "sociable" for the purposes of this research.

Most interesting for the purposes of this study, however, was the evidence of "semantic drift" between topics oriented towards the content, and those derived from the interactional context[48].

Observing the constant tying and switching between different topics and orientations of reference to content, context, or other media objects collected and "heckled" onto the screen by participants suggests that any attempts at "mapping" formal media ontologies to ontologies of practice are likely to stymie this existing communicative process by imposing a-priori topic boundaries, or more likely, become irrelevant to, or provide further fodder for this conversational process of "semantic drift".

Alongside the amenability of the data to conversation analysis, this evidence of "semantic drift" could be seen as another marker of conversation that helps constitute a working concept of "sociability"[49].

8.2 Future Work

Although the results of this study indicate a lack of correspondence between formal and pragmatic ontologies of content in communication, there are still

[48] An attempt to use Dynamic Topic Analysis (Herring, 2003), to analyse the "semantic distance" between heckles was aborted for this reason, see appendix B.

[49] The provenance of the Heckle system in *The People Speak's* development of its first prototype was intended to enable this kind of tangential, free-flowing discussion, see appendix E.

intriguing possibilities for “conversational annotation”.

The most immediate possibility using existing research presented here is to perform a Conversation Analysis on the heckle data, to explore the detail of the interactions fully, and gain further insight into the structural relationships between conversational and formal ontologies.

Collaborative research with Toby Harris of the BBC Stories team on the possibility of integrating the two approaches to media content annotation is already planned, investigating the degree to which conversational annotation tools such as Heckle might integrate and use structured representations of content to present enhanced annotation interfaces to users.

For example, a heckle interface that has an underlying model of the characters, props or narrative elements currently on the screen might be able to offer clickable icons with which viewers could initiate annotations, automatically providing explicit references to content, or links between content and contextual markers.

A second version of the Heckle tool was developed in the course of this research (see Appedix **??**), based on notes and feedback from the user tests. The possibility to have far larger user groups, or non co-located groups use the tool will enable larger data sets to be assembled for subsequent research, and for a tighter integration of data collection and analysis processes. A prominent response from users was also the desire to be able to reply to heckles explicitly, using a “reply” button. Heckle 2 is designed to facilitate this kind of interaction, providing more auditable data about sequences and conversational structures within the data.

From a “use case” perspective, approaches derived from, but qualitatively very different to, “Human Computation”[50] (von Ahn, 2007), might make use of the observed “semantic drift” in topical reference and relatedness in conversational annotation to provide a degree of conversationally-derived serendipity to recommendation engines; a factor that is acknowledged to be perceived as

[50] See appendix D

beneficial but difficult to emulate (Herlocker et al., 2004), (Cremonesi & Turrin, 2010).

Building on the capabilities of Human Computation and content-analysis approaches to the "low level" semantics of image and video labelling, multi-modal annotations via systems such as Heckle could multiply the richness of conversational content annotations by layering all the labels of each heckled video or image onto the annotated segment of video. If segments of pre-annotated video are themselves used as "heckles", iteratively applied to video during multiple "viewing events", the layering effect is further multiplied.

The potential qualitative differences of this kind of conversationally annotated video archive from a structured, broadcaster-centric model are intriguing. Although the archive video content itself may stay the same, its interpretation, captured via the conversations and interactional contexts in which it is used are likely to change over time. A contemporary video might, over time, become nostalgic, or used ironically to illustrate some future event or significance which would be impossible to predict at the time of production.

Overall, these use cases illustrate the central findings of this research: that systems designed to leverage and apply structured data about TV narratives to viewer interaction, by means of Social TV or "second screen" devices[51] should be aware of the different understandings of and attitudes towards "content" evident in broadcaster-supplied metadata and industry-centric approaches, and those implied by the conversational uses of Social TV.

[51] A recent rationale for the provision of augmented TV services via "companion" or "second screen" devices was proposed by Kramskoy (2011) at the BBC, using the term "orchestrated media" to describe the synchronised provision of auxiliary programme information, quizzes, competitions or other opportunities for viewers to "inform, educate and entertain" themselves while watching TV.

References

Abreu, J., Almeida, P., & Branco, V. (2002). 2BeOn-Interactive television supporting interpersonal communication. In *Multimedia 2001: proceedings of the Eurographics Workshop in Manchester, United Kingdom, September 8-9, 2001*, (pp. 199–208). Springer Verlag Wien.

Ang, I. (1996). *Living room wars: Rethinking media audiences for a postmodern world*. Media and communication studies: Cultural studies. London: Burns & Oates.

Aroyo, L., Kaptein, A., Palmisano, D., Conconi, A., Nixon, L., Vignaroli, L., Dietze, S., Nufer, C., & Yankova, M. (2009). NoTubemaking TV a medium for personalized interaction. In *EuroITV 2009 Networked Television*. EuroITV.

Aubert, O., & Prié, Y. (2005). Advene: active reading through hypervideo. In *Proceedings of the sixteenth ACM conference on Hypertext and hypermedia*, (pp. 235–244). ACM.

Baca, M. (2008). *Television meets Facebook : Social Networking Through Consumer Electronics*. Msc, Massachusetts Institute of Technology.

Barkhuus, L. (2009). Television on the internet: new practices, new viewers. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, (pp. 2479–2488). ACM.

Bello, J., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. (2005). A tutorial on onset detection in music signals. *Speech and Audio Processing, IEEE Transactions on*, *13*(5), 1035–1047.

Berners-Lee, T. (2006). Notation3 (N3) A readable RDF syntax. URL `http://www.w3.org/DesignIssues/Notation3`

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 34–43.

Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., & Tscheligi, M. (2008). Trends in the living room and beyond: results from ethnographic studies using creative and playful probing. *Computers in Entertainment (CIE)*, *6*(1), 5.

Brickley, D., & Miller, L. (2005). FOAF Vocabulary Specification. URL `http://xmlns.com/foaf/0.1/`

Brown, J. (2002). *Shakespeare and the theatrical event*. Palgrave Macmillan.

Brunelli, R., Mich, O., & Modena, C. (1996). A survey on video indexing. *Journal of Visual Communications and Image Representation*, *10*, 78–112.

Butsch, R. (2003). Popular Communication Audiences: A Historical Research Agenda. *Popular Communication*, *1*(1), 15–21.

Buzzard, K. (2002). The peoplemeter wars: A case study of technological innovation and diffusion in the ratings industry. *Journal of Media Economics*, *15*(4), 273–291.

Campbell, D., Yonish, S., & Putnam, R. (1999). Tuning in, tuning out revisited: A closer look at the causal links between television and social capital. In *Annual Meeting of the American Political Science Association, Atlanta*.

Cattelan, R. G., Teixeira, C., Goularte, R., & Pimentel, M. D. G. C. (2008). Watch-and-comment as a paradigm toward ubiquitous interactive video editing. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *4*(4), 1–24.

Cesar, P., Bulterman, D., Geerts, D., Jansen, J., Knoche, H., & Seager, W. (2008a). Enhancing social sharing of videos: fragment, annotate, enrich, and share. In *Proceeding of the 16th ACM international conference on Multimedia*, (pp. 11–20). ACM.

Cesar, P., Bulterman, D., & Jansen, A. (2006). The ambulant annotator: empowering viewer-side enrichment of multimedia content. In *Proceedings of the 2006 ACM symposium on Document engineering*, (pp. 186–187). ACM.

Cesar, P., Bulterman, D., & Jansen, A. (2008b). Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content. In *Proceedings of EuroITV*, (pp. 168–177). Springer.

Cesar, P., Bulterman, D. C. a., Jansen, J., Geerts, D., Knoche, H., & Seager, W. (2009). Fragment, tag, enrich, and send. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *5*(3), 1–27.

Cesar, P., & Chorianopoulos, K. (2007). The Evolution of TV Systems, Content, and Users Toward Interactivity. *Foundations and Trends in Human-Computer Interaction*, *2*(4), 373–95.

Cesar, P., Chorianopoulos, K., & Jensen, J. F. (2008c). Social television and user interaction. *Computers in Entertainment*, *6*(1), 1.

Cesar, P., & Geerts, D. (2011). Past, present, and future of social TV: A categorization. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, (pp. 347–351). IEEE.

Chorianopoulos, K. (2007). Content-Enriched Communication-Supporting the Social Uses of TV. *J.Communications Network*, *6*(1), 23.

Cisco (2010). Next Generation Video: Industry Trends and UK Customer Research. Tech. rep., Cisco Systems Inc., Ipswich.

Clark, H., & Brennan, S. (1991). Grounding in communication. *Perspectives on socially shared cognition*, *13*(1991), 127–149.

Colaco, A., & Kim, I. (2010). Back Talk: An auditory environment for sociable television viewing. *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*.

Coppens, T., & Trappeniers, L. (2004). AmigoTV: towards a social TV experience. In *Proceedings of EuroITV*. Brighton: University of Brighton, UK.

Cremonesi, P., & Turrin, R. (2010). Time-evolution of IPTV recommender systems. *Proceedings of the 8th international interactive conference on Interactive TV&Video - EuroITV '10*, (p. 105).

Davis, M. (2000). Media Streams: an iconic visual language for video annotation. *Proceedings 1993 IEEE Symposium on Visual Languages*, (pp. 196–202).

Diakopoulos, N. (2009). *Collaborative annotation, analysis, and presentation interfaces for digital video*. Ph.d., Georgia Institute of Technology.

Diakopoulos, N., Luther, K., & Essa, I. (2008). Audio Puzzler: piecing together time-stamped speech transcripts with a puzzle game. In *Proceeding of the 16th ACM international conference on Multimedia*, (pp. 865–868). ACM.

Fagá Jr, R., Motti, V., Cattelan, R., Teixeira, C., & Pimentel, M. (2010). A social approach to authoring media annotations. In *Proceedings of the 10th ACM symposium on Document engineering*, (pp. 17–26). ACM.

Geerts, D. (2006). Comparing voice chat and text chat in a communication tool for interactive television. *Proceedings of the 4th Nordic conference on Human-computer interaction changing roles - NordiCHI '06*, (pp. 461–464).

Geerts, D., & De Grooff, D. (2009). Supporting the social uses of television: sociability heuristics for social TV. In *Proceedings of the 27th international conference on Human factors in computing systems*, (pp. 595–604). Boston, MA, USA.

Gligorov, R., Hildebrand, M., van Ossenbruggen, J., Schreiber, G., & Aroyo, L. (2011). On the role of user-generated metadata in audio visual collections. In *Proceedings of the sixth international conference on Knowledge capture*, (pp. 145–152). ACM.

Goffman, E. (1966). *Behavior in public places: notes on the social organization of gatherings*. Free press paperback. New York, NY, USA: Free Press.

Goldmedia (2010). EPGs and TV Middleware Applications : Market assessment and forecasts to 2014.

Hansen, M. (1994). *Babel and Babylon: spectatorship in American silent film*. Harvard University Press.

Harboe, G. (2009). In search of social television. In P. Cesar, D. Geerts, & K. Chorianopoulos (Eds.) *Interactive Television: Immersive Experiences and Perspectives*, chap. 1, (pp. 1–13). IGI Global.

Harboe, G., Massey, N., Metcalf, C., Wheatley, D., & Romano, G. (2007). Perceptions of value: The uses of social television. *Interactive TV: a Shared Experience*, (pp. 116–125).

Harboe, G., Massey, N., Metcalf, C., Wheatley, D., & Romano, G. (2008a). The uses of social television. *Computers in Entertainment*, *6*(1), 1.

Harboe, G., Metcalf, C., Bentley, F., Tullio, J., Massey, N., & Romano, G. (2008b). Ambient social tv: drawing people into a shared experience. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, (pp. 1–10). ACM.

Harper, R., Regan, T., & Rouncefield, M. (2006). Taking hold of TV: learning from the literature. In *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, (pp. 79–86). ACM.

Harris, M. T. (2010). *Programme Project Report 2010 Digital Narratives at the BBC*. Msc, Queen Mary University, London.

Harrison, S. (1996). Re-P1ace-ing Space: The Roles of Place Collaborative Systems. *Citeseer*, *7*, 67–76.

Hartley, J. (1999). *Uses of television*. London: Routledge.

Harvey, A. S. (1990). Time Use Studies for Leisure Analysis. *Social Indicators Research*, *23*(4), 309–336.

He, L., Sanocki, E., Gupta, A., & Grudin, J. (1999). Auto-summarization of audio-video presentations. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, (pp. 489–498). New York, New York, USA: ACM.

Healey, P. G. T., White, G., Eshghi, A., Reeves, A. J., & Light, A. (2007). Communication Spaces. *Computer Supported Cooperative Work (CSCW)*, *17*(2-3), 169–193.

Herlocker, J. L., Konstan, J. a., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22*(1), 5–53.

Herring, S. (2003). Dynamic topic analysis of synchronous chat. In *New Research for New Media: Innovative Research Methodologies Symposium Working Papers and Readings.*. Minneapolis, MN: University of Minnesota School of Journalism and Mass Communication.

Herring, S., & Kurtz, A. (2006). Visualizing dynamic topic analysis. In *Proceedings of CHI'06*, (pp. 1–6). ACM Press.

Hoggart, R. (1957). *The Uses of Literacy*. No. 11 November in Classics in communication and mass culture series. Transaction Publishers.

Honey, C., & Herring, S. (2009). Beyond microblogging: Conversation and collaboration via Twitter. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, (pp. 1–10). IEEE.

Hunter, J. (2001). Adding multimedia to the semantic web-building an mpeg-7 ontology. In *Proc. of the First International Semantic Web Working Symposium (SWWS01)*, D, (pp. 261–283). Stanford, CA: Citeseer.

Ito, M., & Okabe, D. (2005). Technosocial situations: Emergent structurings of mobile email use. In M. Ito, & D. Okabe (Eds.) *Personal, portable, pedestrian: Mobile phones in Japanese life*, (pp. 257–273). Cambridge, MA: MIT Press.

Jameson, F. (1998). *The cultural turn*. Verso London.

Jewell, M., Lawrence, K., Tuffield, M., Prugel-Bennett, A., Millard, D., Nixon, M., & Shadbolt, N. (2005). OntoMedia: An ontology for the representation of heterogeneous media. In *In proceedings of Multimedia Information Retrieval Workshop 2005 (MMIR 2005)*. Salvador, Brazil: ACM SIGIR.

Johansen, R. (1988). *GroupWare: Computer Support for Business Teams*. New York, NY, USA: The Free Press.

Klym, N., & Montpetit, M. (2008). Innovation at the edge: Social TV and beyond.

Koenen, R. (2002). MPEG-4 Overview - (V.21 Jeju Version).

Kramskoy, J. (2011). BBC - Research and Development: Orchestrated Media - beyond second and third screen (II).
URL `http://bbc.in/qgagy2`

Lawrence, F., Tuffield, M., Jewell, M., Prügel-Bennett, A., Millard, D., Nixon, M., Schraefel, M., & Shadbolt, N. (2005). Ontomedia-creating an ontology for marking up the contents of heterogeneous media. In *Proceedings of Ontology Patterns for the Semantic Web ISWC-05 Workshop*, (pp. 1–15). Galway, Ireland.

Liu, Y., Shafton, P., Shamma, D., & Yang, J. (2007). Zync: the design of synchronized video sharing. In *Proceedings of the 2007 conference on Designing for User eXperiences*, (pp. 1–8). ACM.

Lull, J. (1980). The Social Uses of Television. *Human Communication Research*, *6*(3), 197–209.

Luyten, K., Thys, K., Huypens, S., & Coninx, K. (2006). Telebuddies: social stitching with interactive television. In *CHI'06 extended abstracts on Human factors in computing systems*, (pp. 1049–1054). ACM.

Martínez, J. M. (2004). MPEG-7 Overview (version 10).

McParland, A. (2002). TV-Anytime - using all that extra data. *R&D White Paper, BBC*.

Melville, P., Mooney, R. J., & Nagarajan, R. (2002). Content-Boosted Collaborative Filtering for Improved Recommendations. In *Proceedings of the 18th National Conference on Artificial Intelligence*, vol. AAAI-2002, (pp. 187–192).

Meyrowitz, J. (1985). *No sense of place: the impact of electronic media on social behavior*. Oxford paperbacks. Oxford University Press.

Mittell, J. (2001). A Cultural Approach to Television Genre Theory. *Cinema Journal*, *40*(3), 3–24.

Montpetit, M., Klym, N., & Mirlacher, T. (2010). The future of IPTV. *Springer Multimedia Tools and Application Journal*, *53*(3), 519–532.

Motti, V. G., Fagá, R., Catellan, R. G., Pimentel, M. D. G. C., & a.C. Teixeira, C. (2009). Collaborative synchronous video annotation via the watch-and-comment paradigm. *Proceedings of the seventh european conference on European interactive television conference - EuroITV '09*, (p. 67).

Nack, F., van Ossenbruggen, J., & Hardman, L. (2005). That obscure object of desire: multimedia metadata on the Web, part 2. *IEEE Multimedia*, *12*(1), 54–63.

Nathan, M., Harrison, C., Yarosh, S., Terveen, L., Stead, L., & Amento, B. (2008). CollaboraTV: making television viewing social again. In *Proceeding of the 1st international conference on Designing interactive user experiences for TV and video*, (pp. 85–94). ACM.

Oehlberg, L., Ducheneaut, N., Thornton, J., Moore, R., & Nickell, E. (2006). Social TV: Designing for distributed, sociable television viewing. In *Proc. EuroITV*, vol. 2006, (pp. 25–26).

Oumard, M., Mirza, D., Kroy, J., & Chorianopoulos, K. (2008). A cultural probes study on video sharing and social communication on the internet. *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts - DIMEA '08*, (p. 142).

Petridis, K., Bloehdorn, S., Saathoff, C., Simou, N., Dasiopoulou, S., Tzouvaras, V., Handschuh, S., Avrithis, Y., Kompatsiaris, Y., & Staab, S. (2006). Knowledge representation and semantic annotation of multimedia content. *IEE Proceedings of Vision, Image and Signal Processing*, *153*(3), 255–262.

Pimentel, M. G., Goularte, R., Cattelan, R. G., Santos, F. S., & Teixeira, C. (2007). Enhancing Multimodal Annotations with Pen-Based Information. *Ninth IEEE International Symposium on Multimedia Workshops (ISMW 2007)*, (pp. 207–213).

Pold, S. r., & Andersen, C. (2011). The Scripted Spaces of Urban Ubiquitous Computing: The experience, poetics, and politics of public scripted space. *Fibreculture Journal*.

Putnam, R. D. (1995). Tuning In, Tuning Out: The Strange Disappearance of Social Capital in America. *PS: Political Science and Politics*, *28*(4), 664.

Putnam, R. D. (2001). *Bowling alone: the collapse and revival of American community*. Simon & Schuster.

Rice, M., & Alm, N. (2007). Sociable TV: Exploring user-led interaction design for older adults. *Interactive TV: a Shared Experience*, (pp. 126–135).

Rivest, R., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, *21*(2), 120–126.

Robertson, S., Vojnovic, M., & Weber, I. (2009). Rethinking the ESP game. *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*, (p. 3937).

Sacks, H. (1992). *Lectures on conversation*, vol. 1. Blackwell.

Sacks, H. (1995). Tying techniques. In *Lectures on Conversation*, chap. 11, (pp. 716–724). Wiley-Blackwell.

Schatz, R., Wagner, S., Egger, S., & Jordan, N. (2007). Mobile TV Becomes Social - Integrating Content with Communications. *2007 29th International Conference on Information Technology Interfaces*, (pp. 263–270).

Schegloff, E. A. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *American Journal of Sociology*, *97*(5), 1295.

Schopman, B., Brickly, D., Aroyo, L., Van Aart, C., Buser, V., Siebes, R., Nixon, L., Miller, L., Malaise, V., Minno, M., & Others (2010). NoTube: making the Web part of personalised TV. In *Proceedings of the WebSci10: Extending the Frontiers of Society Online*, (pp. 1–8).

Shamma, D. (2010). Beyond freebird. *XRDS: Crossroads, The ACM Magazine for Students*, *17*(2), 36–38.

Shamma, D., Kennedy, L., & Churchill, E. (2009). Tweet the debates: understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM Workshop on Social Media*, (pp. 3–10). ACM.

Shamma, D., Shaw, R., Shafton, P., & Liu, Y. (2007). Watch what I watch: using community activity to understand content. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, (pp. 275–284). ACM.

Shannon, C. E. (1948). The mathematical theory of communication. 1963. *M.D. computing : computers in medical practice*, *14*(4), 306–17.

Silverstone, R., & Morley, D. (1990). Domestic communication - technologies and meanings. *Media Culture and Society*, *12*(1), 31–55.

Siorpaes, K., & Hepp, M. (2008). Games with a Purpose for the Semantic Web. *IEEE Intelligent Systems*, *23*(3), 50–60.

Smeulders, A., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(12), 1349–1380.

Stamou, G., van Ossenbruggen, J., Pan, J., & Schreiber, G. (2006). Multimedia annotations on the semantic web. *IEEE Multimedia*, *13*(1), 86–90.

Strong, J. (2008). Doctor Who, Series 4, Episode 1, "Partners in Crime".

Svensson, M., & Sokoler, T. (2008). Ticket-to-talk-television: designing for the circumstantial nature of everyday social interaction. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, (pp. 334–343). ACM.

Thaler, S., Siorpaes, K., Simperl, E., & Hofer, C. (2011). A survey on games for knowledge acquisition. Tech. rep., Semantic Technology Institute, Innsbruck.

Tuffield, M., Millard, D., & Shadbolt, N. (2006). Ontological approaches to modelling narrative. In *Proceedings of 2nd AKT DTA Symposium*. Citeseer.

Tullio, J., Harboe, G., & Massey, N. (2008). Investigating the use of voice and text chat in a social television system. In *Proc. EuroITV: Changing Television Environments*, (pp. 163–167). Berlin: Springer.

Van Aart, C., Aroyo, L., Raimond, Y., Brickley, D., Schreiber, G., Minno, M., Miller, L., Palmisano, D., Mostarda, M., Siebes, R., & Others (2009). The NoTube Beancounter: aggregating user data for television programme recommendation. In *Proceedings of the Linked Data on the Web Workshop (LDOW 2009)*, (pp. 1–12). Madrid, Spain: Citeseer.

Van Ossenbruggen, J., Nack, F., & Hardman, L. (2004). That obscure object of desire: multimedia metadata on the web, part-1. *Multimedia, IEEE*, *11*(4), 38–48.

von Ahn, L. (2007). Human computation. *K-CAP '07 Proceedings of the 4th international conference on Knowledge capture*, (pp. 418–419).

von Ahn, L., Blum, M., & Langford, J. (2002). Telling Humans and Computers Apart (automatically): Or how Lazy Cryptographers Do AI. Tech. rep., Carnegie Mellon University, Pittsburgh, PA.

von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 319–326). Vienna, Austria: ACM.

von Ahn, L., & Dabbish, L. (2008). Designing games with a purpose. *Communications of the ACM*, *51*(8), 57.

Weisz, J. D., & Kiesler, S. (2008). How text and audio chat change the online video experience. *Proceeding of the 1st international conference on Designing interactive user experiences for TV and video - uxtv '08*, (p. 9).

Weisz, J. D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R. E., & Konstan, J. A. (2007). Watching Together : Integrating Text Chat with Video. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 877–886). San Jose, California, USA: ACM.

Yew, J., Shamma, D., & Churchill, E. (2011). Knowing funny: genre perception and categorization in social video sharing. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, (pp. 297–306). ACM.

Yu, Z., Zhou, X., Hao, Y., & Gu, J. (2006). TV Program Recommendation for Multiple Viewers Based on user Profile Merging. *User Modeling and User-Adapted Interaction*, *16*(1), 63–82.

Zillmann, D., & Bryant, J. (1985). *Selective exposure to communication*. Communication (Hillsdale, N.J.). L. Erlbaum Associates.

A The Cultural Uses of Social TV Research

This appendix briefly outlines a rationale for the engagement with Social TV as a research context, and an attempt to contextualise it in a broader understanding of television..

Cesar and Geerts' 2011 Social TV "categorisation" attempts to update previous analytic surveys of Social TV (Chorianopoulos, 2007), (Harboe, 2009) to match the "broadness and complexity" of "a semi-chaotic situation which research, industry, and entrepreneurs are still trying to fully understand" (Cesar & Geerts, 2011). They illustrate this situation with an exhaustive survey of the diffusion of Social Network and media sharing technologies via IPTVs, mobile devices and 'Over The Top'[52] (OTT) TV services all labelled "Social TV", and then attempt to clarify it by developing a framework for categorising various Social TV activities[53] such as "Content Selection and Sharing" or "Communication", along with various "aspects" of each category such as "modality" and "presence"[54].

However, as Gunnar Harboe notes at the end of his account of the history, definitions and dimensions of Social TV: "just as convergence is turning Social TV into reality, it may also be rendering the notion obsolete[. . .] social television as a separate concept might no longer make sense." (Harboe, 2009). Rather than searching for a definition of Social TV, the title of a 2008 paper by Harboe, Massey, Metcalf, Wheatley and Romano asks instead about "The Uses of Social Television" (Harboe et al., 2008a). This paper, which sets out to address the lack of studies "in natural environments", that examine "the actual use of social television applications" appropriates and re-organises the title of James

[52] Over The Top services use internet connectivity and software to replace hardware switches and infrastructure and integrated contractual relationships with customers. Typically this term is being used to refer to cloud-based services such as Netflix which provide video on demand service without a connection or a provider contract.

[53] Including "Content Selection and Sharing", "Communication", "Community building", and "Status Update".

[54] Including "Device/network" (hardware/software infrastructure), "Modality" (whether it uses audio/text/video mediation), "Presence" (how other viewers are represented), "Synchronisation"(whether communications are synchronous or asynchronous) and "Strength tie"(the intended scope of the communication i.e. "public" or "family and friends only".

Lull's often-cited 1980 essay "The Social Uses of Television" (Lull, 1980).

This appropriation, as the title of a paper which questions a technological definition of Social TV suggests a "cultural turn" (Jameson, 1998) to the study of Social TV. John Heartley's 1999 book "Uses of Television" argues that although "it is not possible to imagine television as a single object of study", an analytical approach can ask an historical question: "what is television for? What are the uses of television?"(Hartley, 1999). In turn, Hartley's title and research question is an homage to Richard Hoggart's seminal Cultural Studies text "The Uses of Literacy"(Hoggart, 1957), which pioneered the critical analysis of popular pulp fiction as not only a literary study, but also a lens through which to explore the culture of people who use popular literature on their own terms, and to observe how they deploy it in their lives and relationships.

Taking a similar cultural turn, this research project is both critical of, but also inspired by the way the Social TV research literature omits a clear notion of "The Social" and "TV", and concentrates on pragmatically converging technologies, cultural contexts, social practices, and methods of measurement and analysis in novel ways, while retaining the seemingly familiar context of "TV" for evaluation by subjects in field tests. However, in order to use the Social TV literature to derive methods and design tests for this research, a more grounded theoretical framework that takes into account both social and technical aspects of an evolving "situation" in front of the TV is required.

B Dynamic Topic Analysis

Dynamic Topic Analysis (Herring, 2003) (DTA) was applied to the sections of heckle data analysed in part 6.4.5. DTA is a method for analysing the conversational coherence of Computer Mediated Communications (CMC). Researchers code data based on its relationship to prior turns with labels such as "T" (on Topic), "B" (Break), "M" (Meta), or "P" (Parallel), and it's "semantic distance" in terms of reference to an "originary" turn.

However, the results of the “Conversational markers” analysis, and evidence of the “semantic drift” of topic between context and content-relatedness cast doubt on the validity of this method for analysing data gathered in the hybrid live/CMC context of heckled co-present TV viewing, and the typing structure seemed unhelpful to the rest of the analysis (see figure 24).

The visualisations (see figures 22 and 23), produced using VisualDTA software (Herring & Kurtz, 2006) is a useful indication of the relative coherence of heckled conversations.

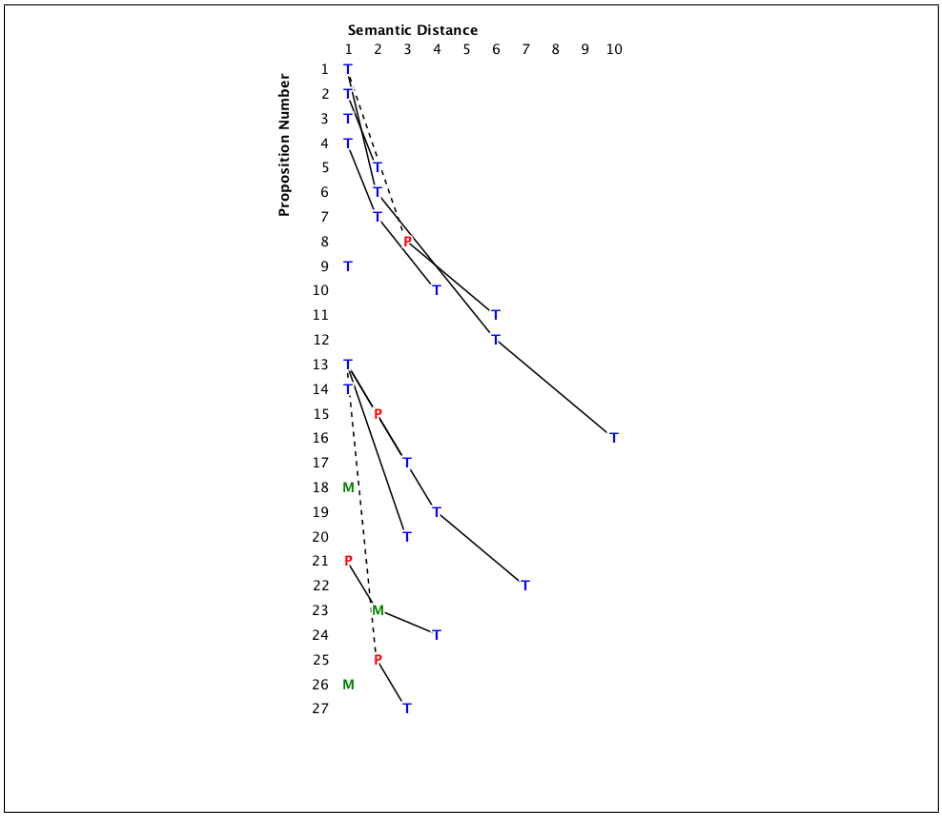

Figure 22: A VisualDTA visualisation of a DTA analysis of the heckle data analysed in figure 18.

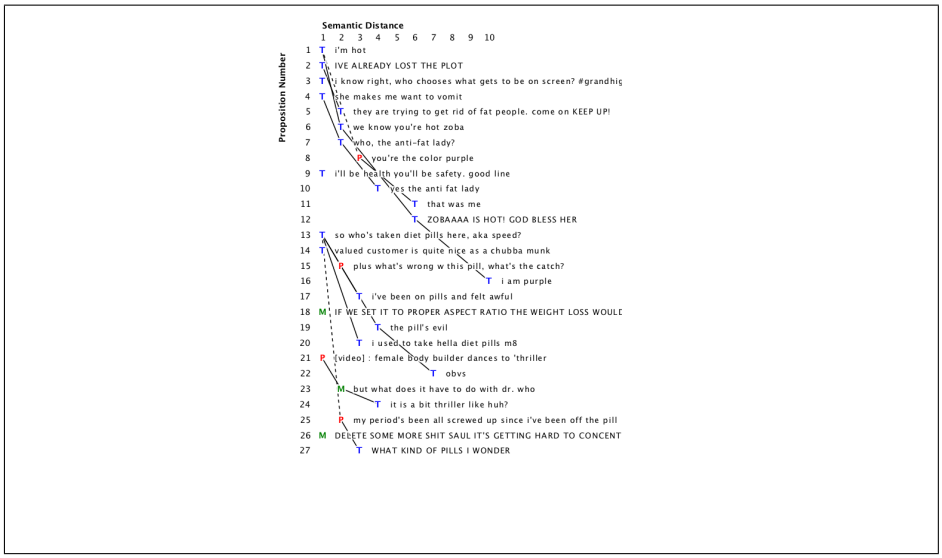

Figure 23: A VisualDTA visualisation of a DTA analysis of the heckle sequence analysed in figure 18.

Proposition	Speaker	Responds to	Relation Type	Distance	Dotted line?	Text
1	lightblue	NA	T	0	0	i'm hot
2	orange	NA	T	0	0	IVE ALREADY LOST THE PLOT
3	grey	NA	T	0	0	j know right, who chooses what gets to be on screen? #grandhighpooba
4	green	NA	T	0	0	she makes me want to vomit
5	pink	2	T	1	0	they are trying to get rid of fat people. come on KEEP UP!
6	purple	1	T	1	0	we know you're hot zoba
7	grey	4	T	1	0	who, the anti-fat lady?
8	purple	1	P	2	1	you're the color purple
9	grey	NA	T	0	0	i'll be health you'll be safety. good line
10	green	7	T	2	0	yes the anti fat lady
11	lightblue	8	T	3	0	that was me
12	orange	6	T	4	0	ZOBAAAA IS HOT! GOD BLESS HER
13	darkblue	NA	T	0	0	so who's taken diet pills here, aka speed?
14	grey	NA	T	0	0	valued customer is quite nice as a chubba munk
15	grey	13	P	1	0	plus what's wrong w this pill, what's the catch?
16	purple	12	T	4	0	i am purple
17	lightblue	13	T	2	0	i've been on pills and felt awful
18	orange	NA	M	0	0	IF WE SET IT TO PROPER ASPECT RATIO THE WEIGHT LOSS WOULDBE M
19	darkblue	15	T	2	0	the pill's evil
20	grey	17	T	3	0	i used to take hella diet pills m8
21	cyan	15	T	0	0	[video] : female body builder dances to 'thriller'
22	red	19	T	3	0	obvs
23	darkblue	21	M	1	0	but what does it have to do with dr. who
24	purple	23	T	2	0	it is a bit thriller like huh?
25	darkblue	17	P	1	1	my period's been all screwed up since i've been off the pill
26	grey	NA	M	0	0	DELETE SOME MORE SHIT SAUL IT'S GETTING HARD TO CONCENTRATE
27	orange	17	T	3	0	WHAT KIND OF PILLS I WONDER

Figure 24: The analysis data for the VisualDTA visualisation of the heckle sequence analysed in figure 18.

C Social TV annotation tools

This appendix outlines a more detailed rationale for the development of a specialised tool to collect "conversational metadata". A detailed survey of these tools was required to establish the purpose of building a new one.

There is a wealth of research in CSCW and a large number of domain-specific commercial products for the human-assisted production of multimedia metadata, designed to help overcome this semantic gap in indexing and cataloging multimedia assets such as medical, military and research corpora. Given that Social TV has been characterised (Chorianopoulos, 2007) in terms of CSCW's time-space matrix (Johansen, 1988)[55], these tools would seem to be pragmatically compatible. However Social TV researchers have differentiated their work from CSCW approaches by pointing to "limited investigation in the context of leisure activities, such as TV" (Chorianopoulos, 2007, p.24), or "informal TV sociability" (Oumard et al., 2008, p.143), (Motti et al., 2009), (He et al., 1999).

Cesar et al. proposed "viewer-side content enrichment"(Cesar et al., 2006), as an approach, with an "Ambulant Annotator" interface[56] to enable TV viewers to pause, add and share hyperlinks to frames of videos as they watch. Pimentel et al. extended this approach into a "Watch-and-Comment Paradigm" (Pimentel et al., 2007) with an associated Watch-and-Comment Tool (WaCTool) enabling viewers to pause and annotate videos using "digital ink" interfaces and voice notes, processed into text by voice or handwriting recognition and stored as separate XML metadata documents. Later developments from both groups of researchers adopt increasingly "light weight" tools in order to reduce the complexity and cognitive burden of the annotation process by developing a single watching and annotation "second screen" interface (Cesar et al., 2008b), and focusing on the socially-motivated annotation and sharing of media "fragments"

[55]The CSCW time/space matrix plots synchronous and asynchronous modes of communication against collocated and remote communications to analyse the affordances of CSCW scenarios and technologies.

[56]Based on the Ambulant media player http://www.ambulantplayer.org/

(Cesar et al., 2009).

Research on the Collaborative Watch And Comment Tool (CWaCTool[57]) proposed a "Social Approach to Authoring Media Annotations" (Fagá Jr et al., 2010) by adding remote collaboration functions to the existing WaCTool, including a chat function, the ability to use and annotate Youtube[58] videos, and integration with user profiles on Social Networking sites. In their discussion of the chat function, Faga et al. use their findings to refute prior claims that text chat is distracting (Weisz et al., 2007) adding that the users reported enjoying chatting, and were influenced by other users, but because they could pause the video, they "didn't consider the chat feature a problem to focus on video content[*sic*]"(Fagá Jr et al., 2010, p.24) when making their annotations. As Faga et al. do not evaluate the contents of the chat between users, it is not clear whether and to what extent both annotations and communication are grounded in the communication between the users, the video being watched, or whether, as Shamma et al. suggest, that "semantics would be tied to the pragmatics of the annotating application": to the interface and functionality of the CWaCTool itself(Shamma et al., 2009, p.2), (Shamma et al., 2007).

[57] A follow-up to the WaCTool emphasising remote collaboration, see http://code.google.com/p/cwactool/.

[58] http://www.youtube.com

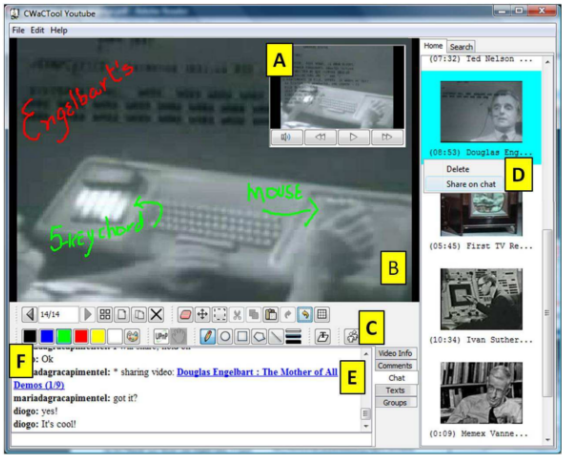

Figure 4: A click on the playback panel minimizes the video (A), places a copy of the corresponding video frame on the ink panel (B) and allows several ink-related annotation options (C). A right-click on the thumbnail (D) allows sharing the video (E) in the chat area (F).

Figure 25: The interface of the CWaCTool

The complexity of the Ambulant Annotator, WaCTool and CWaCTool interfaces, and their reliance on "explicit annotation" (Fagá Jr et al., 2010, p.18) as the central motivation for their use undermines the positioning of these as "leisurely" activities, characterising the situation of using them more as a "lean forward" (Svensson & Sokoler, 2008), (Barkhuus, 2009), (Montpetit et al., 2010), CSCW task-related interaction, than "lean back", leisurely sociability observed between TV viewers (Oehlberg et al., 2006), (Geerts & De Grooff, 2009).

D Games With A Purpose

This appendix explains the concept of Luis Von Ahn's concept of "Human Computation" (von Ahn, 2007) and its approach to harnessing forms of sociability to generate metadata.[59]

Researchers looking for new ways to annotate video that depart from this

[59]The detail of the interactional experience of playing 'Ahn's 'Games With A Purpose" (von Ahn & Dabbish, 2008), on examination, is so different to the experience of using the heckle system, that this section was removed from the thesis. However, it does provide insight into possible uses of this research so is retained in this appendix.

CSCW-like approach to content-enrichment (Diakopoulos, 2009) have pointed to the emerging phenomenon of “Games With A Purpose”(von Ahn & Dabbish, 2008) (GWAP), a method of deriving human-interpreted semantics about multimedia content by involving users in game scenarios, as a potential source of lowering the cost and usability “friction” (Siorpaes & Hepp, 2008) of generating video metadata.

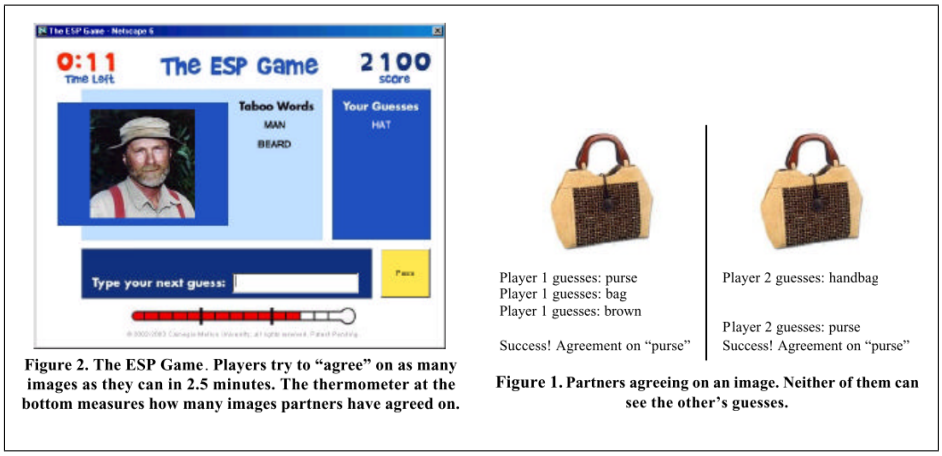

Figure 2. The ESP Game. Players try to “agree” on as many images as they can in 2.5 minutes. The thermometer at the bottom measures how many images partners have agreed on.

Figure 1. Partners agreeing on an image. Neither of them can see the other’s guesses.

Figure 26: The ESPGame interface (von Ahn, 2007)

Luis von Ahn’s thesis on “Human Computation” (von Ahn, 2007) describes the use of everyday human activities, for example, filling in a web “CAPTCHA” form [60] from the reCAPTCHA web service[61] as a means of transcribing words that automated techniques have failed to recognise in large scale book transcription projects. Moving further from the field of computer security and cryptography, von Ahn went on to develop the ESP Game[62], a game in which two remotely located players try to guess at which words the other player is using to annotate an image. When a match is found, the game moves on, and the matched word is then used to annotate an image corpus (von Ahn & Dabbish, 2004).

[60] CAPTCHA stands for “Completely Automated Public Turing test to tell Computers and Humans Apart” (von Ahn et al., 2002), and was developed as a security measure to tell humans apart from spam bots using a reverse Turing test where a computer challenges a user to interpret an image that is distorted enough to confuse CV techniques, and waits for a successful response from a human agent before providing authentication.

[61] http://www.google.com/recaptcha

[62] http://www.gwap.com/gwap/gamesPreview/espgame/

The willingness of large numbers of people to play this game and label huge numbers of images as a byproduct[63] has inspired a surge of research into so-called "Games With A Purpose"[64], as a method of incentivising the production of semantic data through distributed mass leisure activities[65].

Adapting this approach for annotating video and media content Diakopoulos, Luther and Essa describe the AudioPuzzler[66] game (Diakopoulos et al., 2008), which attempts to produce time-stamped audio transcripts from videos by breaking up the audio into overlapping "puzzle pieces", and creating a game scenario in which participants must first transcribe, then piece together a complete transcript.

A number of video-tagging games analogous in function to the ESPGame have since been developed, including von Ahn's PopVideo game [67] and Yahoo's Videotaggame[68], both of which label short video clips, and Waidsa[69], which uses full-length TV shows. All these games choose which annotations to apply to which segments of video based on consensus being achieved between game players about labels applied at roughly the same time. (Thaler et al., 2011).

The use of GWAP has been shown to enable the creation of accurate video transcriptions in the AudioPuzzler project (Diakopoulos et al., 2008), and user-generated consensual TV content annotations from the Waisda project have recently been shown to complement professionally produced metadata in their descriptions of objects in the audio and video stream (Gligorov et al., 2011).

However, although the interfaces to GWAP seem simpler than the "Watch-and-Comment" annotation systems, and the motivation to engage seems more playful, the highly constrained interactions these games demand suggest that the GWAP approach is no more likely to create the "lean back" social experi-

[63] Google acquired the rights to the game and has been using and modifying it, along with other techniques, to annotate their databases for Google Image Search (Robertson et al., 2009)
[64] http://www.gwap.com
[65] See http://www.insemtives.eu/
[66] http://www.audiopuzzler.com/
[67] http://www.gwap.com/gwap/gamesPreview/popvideo/
[68] http://videotaggame.sandbox.yahoo.com/
[69] http://www.waisda.nl/

ence of sociable TV viewing described earlier. Similarly, the reliance on consensus between players to progress through each "turn" of a game (Siorpaes & Hepp, 2008), and the communication model prescribed for players [70] by the "common ground" of the game mechanics seems to limit the scope of the exchanges captured by the game to the exchange of generalized tag-like semantics(Gligorov et al., 2011). Although people play GWAP for fun rather than professionally, in practice this form of "Human Computation" seems even more task-oriented than CSCW-like content annotation processes and interfaces.

[70] Ahn's early work on cryptographic and security systems suggests a compelling explanation for the origins of his thinking on GWAP. Security and cryptography systems are often illustrated with scenarios involving two characters: Alice and Bob, first used in a famous example from Ron Rivest's seminal paper on RSA Public Key Cryptography (Rivest et al., 1978). In Rivest's scenario, Bob needs to send Alice a message without revealing it to third parties in transmission. They each have a "public key" and a "secret key". Bob combines his secret key with Alice's public key to create a "shared secret". The shared secret is then used to encode (or cryptographically "hash") the message before transmission, which can then be decrypted by Alice, using her "secret key" and Bob's "public key" to arrive at the same "shared secret" without having to share secret keys. The ESPGame uses elements of this design in its game dynamic, in which Bob and Alice can encode and decode conceptualisations of the world using a public key (the message) and their own secret key (presumably identical knowledge-bases about the world). The underlying model of human communication assumed by this process is very similar to Shannon and Weaver's "naïve code model" of human communication (Shannon, 1948).

E Heckle, by The People Speak

Since 2007, art collective *The People Speak* have been working on ways of trying to make over 13 years of their conversational oral history archive public and searchable.

This archive consists primarily of recordings of conversations between people who have met and talked around their "Talkaoke table: a pop-up talk-show invented by Michael Weinkove of *The People Speak* in 1997. It involves a doughnut-shaped table, with a host sitting in the middle on a swivel chair, passing the microphone around to anyone who comes and sits around the edge to talk[71] on street corners, at festivals, schools, or conferences.

Figure 27: The Talkaoke Table in use at the National Theatre, London, 2011 as a means for gathering post-show audience response.

The problem that Heckle 1 was designed to address was figuring out what people are talking about in the mountain of video data collected over this time. All the conversations facilitated by *The People Speak* are spontaneous, off the

[71] See figure 27, and http://talkaoke.com for more information.

cuff, and open to people changing topic at any point. This makes the provision of a thematically searchable archive structured archive very challenging.

This challenge is not unique to this specialised context. As suggested by the observations of "semantic drift" in 8.1.4, conversations, questions and answer sessions, and contexts that involve people interacting with each other seem subject to the same contingencies and tangents of meaning.

The People Speak's development of the Heckle system is intended to create hybrid, contingent topical annotations as an outcome of a live, multi-modal conversational process, in which the audience themselves interject turns, queries and "repair" each others spoken or heckled contributions.

Using "Heckle", an operator, or multiple participants in a conversation may search for and post google images, videos, web links, Wikipedia articles or 140 characters of text, which then appear superimposed on a live, projected video of the conversation[72].

In figure 27, the people sitting around the Talkaoke table are not focused on the screens on which the camera view is projected live. The aim of the Heckle system is not to compete with the live conversation as such, but to be a shared "backchannel", throwing up images, text and contextual explanations on the screen that enable new participants to understand what is going on and join in the conversation.

[72] See figure 27

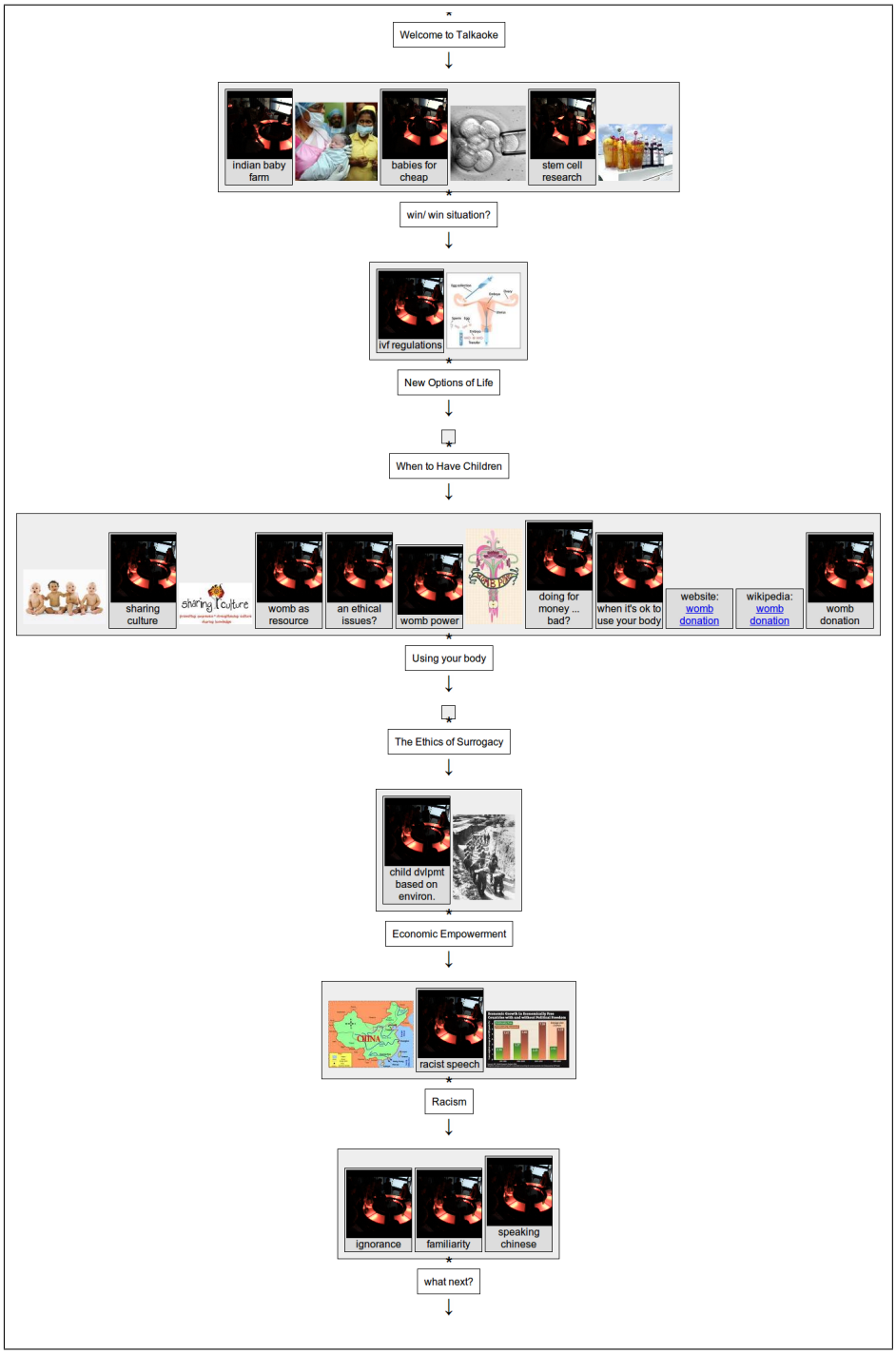

Figure 28: The Heckle system's "cloud" mode, which shows all image, text and video heckles so far, including snapshots from the live video feed taken at the moment that each texts are sent to the screen.

Shown in figure 28, the Heckle system also has a “cloud” mode, in which it displays a linear representation of the entire conversation so far, including snapshots from the video at the moment that a heckle was created, alongside images, keywords, “chapter headings” and video.

This representation of the conversation is often used as part of a rhetorical device by the Talkaoke host to review the conversation so far for the benefit of people who have just sat down to talk. A “Heckle facilitator” can temporarily bring it up on a projection or other nearby display and the host then verbally summarises what has happened so far.

Heckling also often functions as a modifier for what is being said. Someone is talking about a subject, and another participant or viewer posts an image which may contradict or ridicule their statement; someone notices and laughs, the participants’ attention is drawn to the screen momentarily, then returns to the conversation with this new interjection in mind. Some participants use the Heckle system because they are too shy to take the microphone and speak. It may illustrate and reinforce or undermine and satirize. Some heckles are made in reply to another heckle, some in reply to something said aloud, and vice versa.

If keywords are mentioned in the chat, those keywords can be matched to a time-code in the video, in effect, the heckled conversation becomes an index for the video recorded conversation: the conversation annotates the video.