Media of Things: Supporting the Production and Consumption of Object-based Media with the Internet of Things

Gerard Wilkinson

Doctor of Philosophy

Newcastle University & British Broadcasting Corporation

March 2019

Abstract

Visual media consumption habits are in a constant state of flux, predicting which platforms and consumption mediums will succeed and which will fail is a fateful business. Virtual Reality and Augmented Reality could be the 3D TVs that went before them, or they could push forward a new level of content immersion and radically change media production forever. Content producers are constantly trying to adapt to these shifts in habits and respond to new technologies. Smaller independent studios buoyed by their new-found audience penetration through sites like YouTube and Facebook can inherently respond to these emerging technologies faster, not weighed down by the "legacy" many. Broadcasters such as the BBC are keen to evolve their content to respond to the challenges of this new world. Producing content that is both more compelling in terms of immersion, and more responsive to technological advances in terms of input and output mediums. This is where the concept of Object-based Broadcasting was born, content that is responsive to the user consuming their content on a phone over a short period of time whilst also providing an immersive multi-screen experience for a smart home environment.

One of the primary barriers to the development of Object-based Media is in a feasible set of mechanisms to generate supporting assets and adequately exploit the input and output mediums of the modern home. The underlying question here is how we build these experiences, we obviously can't produce content for each of the thousands of combinations of devices and hardware we have available to us. I view this challenge to content makers as one of a distinct lack of descriptive and abstract detail at both ends of the production pipeline. In investigating the contribution that the Internet of Things may have to this space I first look to create well described assets in productions using embedded sensing. Detecting non-visual actions and generating detail not possible from vision alone. I then look to exploit existing datasets from production and consumption environments to gain greater understanding of generated media assets and a means to coordinate input/output in the home. Finally, I investigate the opportunities for rich and expressive interaction with devices and content in the home exploiting favourable characteristics of existing interfaces to construct a compelling control interface to Smart Home devices and Object-based experiences. I resolve that the Internet of Things is vital to the development of Object-based Broadcasting and its wider roll-out.

Publications Arising from Thesis

The following publication arose from Chapter 3, Media of Things:

Gerard Wilkinson, Tom Bartindale, Tom Nappey, Michael Evans, Pete Wright, Patrick Olivier. 2018. **Media of Things: Supporting the Production of Metadata Rich Media Through IoT Sensing.** Proceedings of the *SIGCHI Conference on Human Factors in Computing Systems - CHI '18, ACM Press*.

The following publication arose from Chapter 4, Integration:

Gerard Wilkinson, Dan Jackson, Andrew Garbett, Reuben Kirkham, Kyle Montague. 2020. CryptoCam: Privacy Conscious Open Circuit Television. arXiv.org.

The following publication arose from Chapter 5, Maintaining Control:

Gerard Wilkinson, Ahmed Kharrufa, Jonathan Hook, Bradley Pursglove, Hendrik Haeuser, Nils Hammerla, Gavin Wood, Steve Hodges, Patrick Olivier. 2016. **Expressy: Using a Wrist-worn Inertial Measurement Unit to Add Expressiveness to Touch-based Interactions.** Proceedings of the *SIGCHI Conference on Human Factors in Computing Systems - CHI '16, ACM Press*. http://doi.org/10.1145/2858036.2858223

Project Contribution Summary

This work was conducted in collaboration with a number of organisations and individuals within Open Lab. In the table below, I have highlighted each external contribution to projects in order to properly attribute others works.

Project	External Contributions
Media of Things Production	Tom Nappey was integral to running the production. He designed the kitchen workbench and the developed a means to making the system portable. Tom and I constructed the set in Open Lab, before moving it to the set. BBC R&D organised the production, wrote scripts, auditioned talent and designed the rest of the set.
CryptoCam	CryptoCam software was written by myself and Dan Jackson. The Android application was written by Dr. Kyle Montague.

Acknowledgements

Kyle Montague, Tom Bartindale, Mike Evans & Pete Wright – For all your support and guidance as my supervisors throughout my PhD. The ones who lasted to the end. Special thanks also to Patrick Olivier and Thomas Ploetz.

Nataly Birbeck, Andy Garbett, Dan Jackson, Tom Nappey, Kyle Montague, Ahmed Kharrufa, Ed Jenkins, Lydia Michie, Daniel Welsh, Aare Puussaar & Tom Maskell – you all know why.

My family – Brian, Maureen, Clare & Andrew.

Finally, I would also like to thank our public broadcaster the British Broadcasting Corporation. This work would not have been possible without the funding, support and guidance of BBC Research & Development. I would like to thank all of those whom I have worked with and alongside in MediaCityUK, Salford over the past 4 years. Special mention to Mike Evans, Phil Stenton, Mark Lomas & Dan Hett.

Table of Contents

List of Figures

List of Tables

Introduction

Chapter One

Chapter 1. Introduction

1.1. Industry Pressures

Over the past two decades the practices of producing and consuming visual media have dramatically evolved. At the turn of the millennium, traditional broadcast TV consumption with a set top box or terrestrial TV was the predominant consumption medium [86,108,116]; this was largely due to the highest bandwidth communication medium available to the home being a TV ariel or set top box. A quarter of households in the UK had internet connections but they were often slow and unreliable [71].

During this time Broadcasters focussed their attention on improving picture quality, both in terms of resolution and colour depth, and sound quality through increasing bitrate and channels. Their success in this endeavour lead to increased consumer demand for content and crucially new mechanisms to deliver this content. Alongside the ever-increasing interconnectivity of homes, through proliferation of home broadband connections, broadcasters introduced online on-demand offerings through services like BBC iPlayer[1] and Hulu[2]. Through these mediums they began offering content to a wider range of devices too: laptops, smartphones and others came into the fold for media consumption. In response, visual media researchers and broadcasters are looking at ways of bridging the divide between this broader network of devices in the home and the TV/smartphone/tablet [17,68,93]. This expansion of content consumption mediums led researchers to experiment with more interactive media experiences and taking advantage of the bi-directional communication medium of the Internet and the home [55]. Modern day homes are now vibrant interconnected spaces with a wealth of sensing mediums, connected equipment and appliances. The highest bandwidth communication medium is now Internet connected devices (with higher reliability and presence now at 90% in UK homes [71]). The portability, presence-at-hand and wealth of sensing input and outputs of such devices enable a new level of communication through media, moving beyond these siloed input and output mediums towards a more interconnected and immersive consumption experience for visual media.

---

[1] BBC iPlayer, https://bbc.co.uk/iplayer
[2] Hulu, https://hulu.com

1.2. A Fertile Landscape

Available output and input mediums are now extremely diverse. Broadcasts must similarly diversivy their offering to consumers, look at exploiting these connected spaces and environments, tailoring their media to better suit these devices and spaces – both in terms of physical screen real-estate and other input/output mediums – to create more compelling media experiences.

Using the modern-day living room as an example – a traditional dramatic piece may exploit a single high-resolution screen and perhaps surround sound to create an immersive experience. However, moving beyond this, multiple devices could be exploited in novel ways to further immerse the viewer in the content. For example, consider smart lighting in the home: during a dark or gloomy scene, a consumption experience might dim the lights; whilst during an explosion it could create an ambient orange flash to match the scene. Alternatively, a smartphone in the room could ring in unison with the film or instead of the content being broadcast through the TV, it could be played through the earpiece of the smartphone. Feeding back into the experience, a search of a house could require use of the TV remote as a flash light scouring the room to find a piece of information.

In a didactic context, the system could queue up tasks for a user. In a kitchen, for example, tasks could be: turning on an appliance to pre-heat to the right temperature; preparation of a selection of vegetables with a knife and chopping board. These tasks could be automated, or completion detected to move the film onto the next stage respectively. A player could detect which ingredients or equipment are available to prepare a specific recipe – the media could be selected or adjusted to suit. Film could also be adapted to a user's accessibility needs. For example, this could be done through: substitution of steps; providing content through non-visual mediums; or perhaps just providing more context and information to certain steps.

In constructing these experiences broadcasters are faced with a piecemeal, highly unpredictable environment, with thousands of device screens of varying sizes, capabilities and input and output modalities. Furthermore, the portability of many devices complicates this problem space. In order to approach some sense of universality broadcasters looked to constrain their approach to this form of content, with a conceptual framework for more responsive and adaptive media. This framework is developed with the aim of developing production pipelines and content delivery networks/systems that can support the array of experiences proffered. The industry has termed this conceptualisation Object-based Broadcasting (OBB). This comes from the software engineering term Object-based Programming – encapsulating functionality into classes with a clearly defined purpose, a set of clearly defined inputs, outputs and methods to execute. Media assets can similarly be encapsulated with well described content and purpose, functionality like adaption to fit different screens or highlighting specific areas of the asset for different purposes.

1.3. What is Object-based Broadcasting?

In order to realise the vision of responsive, adaptive and immersive broadcast experience, content makers need to look at the underlying requirements for such experiences, both in terms of the media production itself and the content distribution and playback. Object-based Broadcasting introduces new requirements from developed visual media; films may need to be dynamically reframed or reorganised, or even a completely new perspective retrieved.

Object-based Broadcasting hypothesises an approach to content creation and consumption that enables fundamentally Contextual Reconfiguration of the media. The possibilities for configuration are a combination of the following:

Temporal Configuration – the length of a piece of media can be configured. The media may be extended through relevant elaborations (e.g. adding detailed instructions to a veg prep.) or reduced to only the most salient points (e.g. goals in a football match). The media could even be flattened entirely to a textual representation to be consumed at the users own pace. Crucially this is dynamic, each reconfiguration is not neccesarily fixed (i.e. a short and long representation is not manually produced).

Screen Configuration – content can be adjusted to suit available screen. Irrespective of screen size or aspect ratio content could be resized to fit (e.g. a portrait phone could avoid letterboxing). This re-framing of the media may be performed based upon the saliency of regions of the media.

Environmental Configuration – media may adapt to exploit its environment, providing a more contextually appropriate experience for the playback area. Provided an understanding of the spatial configuration of a playback environment, output may be designed to be more focussed (e.g. a noise from upstairs in a show could be played through a speaker upstairs in a home) or perhaps more detailed (e.g. a supplementary screen showing cookery show steps in text). There may also be a bi-directional dynamism here, the environment may not be static, a user could adjust the space during playback (e.g. lighting or interacting with a device) – the media may respond.

Produced content therefore needs organisation – structured and labelled video and audio assets, in order to develop reconfigurable media experiences. Crucially, this is not necessarily just a single video and audio track; multiple video and audio tracks may be available, with meta-data describing the scene and each track, sensor data from telemetry in the production and any suitable graphics or overlays may also be provided. This structured form of an asset library provides a bedrock for a largely unbounded set of reconfiguration possibilities.

Object-based Broadcasting is the concept of reconfigurable, adaptive and responsive media, the produced output is Object-based Media. I will refer to these two terms throughout this thesis.

1.3.1. Content Delivery and Object-based Broadcasting

Content delivery is the next challenge with regards to reconfigurable media experiences. In many ways, delivery bounds the reconfiguration possibilities, and this is a challenge to existing delivery networks and technologies. Traditional broadcast technology simply does not have the required bandwidth for communication, nor does it provide adequate backchannel interfaces to support the media responding to playback environment events. OBB as a concept does not necessarily provide answers to this problem. Instead it conceptualises a multitude of approaches and each is highly aligned with its relevant consumption experience. At its simplest, content delivery amounts to the level of reconfiguration and responsiveness. Striking a balance by moving content closer to the playback edge, enabling more dynamism in configuration, with more raw assets available to players but at the expense of bandwidth requirements on both the broadcaster and the end user. Conversely, making more final content decisions at the broadcast end (with fewer assets available to the players) provides fewer possibilities for reconfiguration but a more feasible proposition for content delivery solutions.

1.3.2. Immersive Smart Home

The proliferation of Internet of Things and Smart Home technologies has provided an array of new input and output modalities for the home – some are abstract (e.g. colour lighting) and some are complex (e.g. voice interfaces). Content makers are looking to exploit these new interfaces and mediums to create new immersive experiences for users. Naturally, smart home device manufacturers are willing recipients of content, and as such there is a natural synergy for both. As manufacturers strive for deeper integration of devices and services, so too broadcasters look to provide more integrated media playback, exploiting this breadth of input and output mediums available. However, the piecemeal nature of the home has provided a rocky and untamed landscape for content makers to build upon. Object-based Broadcasting conceptualises these issues of potentially millions of combinations of potential input and output devices in the home through a similar approach to productions – well described environments. A device may describe its functionality in such a way that it can be best exploited by a playback experience, viewport dimensions, orientation sensing etc. When considered in conjunction with well described media assets, we have Object-based Broadcasting as a concept. Solving, or mitigating, the aforementioned production and content delivery problem spaces is fundamental to constructing these reconfigurable media experiences that broadcasters wish to explore.

In constructing Object-based experiences there are two key areas that are currently underexplored and underutilised: the integration of content and services between production and consumption – making the required data available; and the development of narrative “engines” and prototypes encapsulating the logic behind such experiences – how the content responds to a device both in terms of output and relevant input, and applying the well described media in a suitable manner. In exploring these two fundamental areas of Object-based Broadcasting I look to realise the opportunities for OBB. Object-based Broadcasting provides a means for us to describe this complex landscape, conceptualising these novel, interactive and adaptive experiences founded in well described media assets and consumption experiences.

1.4. Existing Responsive Experiences

Reconfigurable and responsive media experiences themselves are not necessarily novel, though they may seem to be to the average consumer. Second screen experiences are, as the name suggests, including a secondary display that is synced to the media playback in some way. Typically, these setups provide supplementary information about characters, storylines or perhaps even basic ways to influence the playback. The media assets supporting second screen are usually constructed manually, time-sequenced meta-data providing appropriate information at the required points. Branching narrative is another media experience with limited reconfigurability. Branching here refers to the ability of a user to select a path through the media. Typically, the user is presented with a set of choices at certain stages in the narrative, each decision influencing the subsequent media in some way. Branching narrative experiences have a higher level of structural complexity, with elements of the narrative and media assets used or not used based upon a set of decisions from earlier in the show. However, these media assets are still pre-packaged, with each path decided ahead of production. These branching points are usually manually created with subsequent dependencies carefully crafted to maintain continuity and the narrative. Again, as these are manually constructed with simple logic models backing the experience, their decision arcs are generally limited.

While dependent on budgetary constraints, constructing complex media experiences like second screen and branching narrative are often prohibitively expensive. Manually constructing the branches of the narrative and covering all possibilities can be an arduous task. To avoid ending up being too expensive, such content tends to end up with overly simplistic reconfigurability as a result.

Ad-skipping like TiVo SkipMode[3] provide some basic examples of reconfigurability, using assumed advertisement break times and some basic vision processing. Netflix[4] provides intro and outro skipping. These types of reconfiguration are purely for brevity.

We can also take, as an example, audio descriptions which are generated in post-production, typically from a person watching and listening to a show and relaying the most important information through text. Much of this information is providing context

[3] TiVo SkipMode, https://support.tivo.com/articles/Features_Use/SkipMode
[4] Netflix, https://netflix.com

to a scene, but also in some cases providing more basic scene descriptions such as tracking moving objects, or logging which people are in the scene. This manual labelling is feasible for a standard film, however when introducing it to a reconfigurable or otherwise responsive experience this quickly becomes impractical.

This is where Object-based experiences could offer some answers to these problems of reconfigurability and responsiveness. Rather than bootstrapping existing processes to satisfy these experiences, one can take a step back, and work from a point of needing to maintain as much information about the shoot as possible and consider what needs to change. With asset labelling directly, we can consider the genericism of labelling: “Eggs are whisked in bowl” as an example audio description, could be decomposed into (BOWL & WHISK IN USE, ACTION WHISK). This structure is more deployable as the original message can be reconstructed but can also be used to inform other experiences such as stage detection.

Table 1: Existing Reconfigurable Media Experiences, their data and generation techniques.

Production Type	Data	Hardware	Generation / Curation
Second Screen	Time-coded supplementary information / media assets	Tablet or Smartphone	Manual labelling of relevant sections of media
Branching Narrative	Narrative decision tree, time-coded decision points and arcs	TV Remote or Smartphone	Filming of each story arc, meta-data describing branching locations and decision tree carefully manually constructed
Multi-Screen (Red Button)	Multi-channel video streams with descriptions, supplementary live information	TV Remote	Multiple camera feeds with typically static descriptions manually generated / curated

CAKE Cox et al. [22]	Dependency tree, ingredient labelling, start and end of segments	Tablet or PC	Significant amount of manual labelling, editing for continuity and linking segments

Table 1 surmises existing research into responsive media based upon the meta-data required to back these experiences, the technologies deployed and the means of generation of data and curation of content. Each of the experiences above require data describing the media assets beyond typical subtitling etc., and the experiences are controlled and content presented through a limited range of mediums (tablets, smartphones, TV). The recurring theme I wish to highlight here is the manual nature through which these content experiences are constructed. These time-consuming processes naturally limit the proliferation of this type of content. Balanced against the limited contribution to immersion that these experiences offer broadcasters have largely turned away from the continued development of second screen and branching narrative content. Object-based Broadcasting encapsulates these previous experiences conceptually but builds upon them, providing an abstraction layer that potentially leads to less explicit meta-data needing to be generated. This thesis focusses on this process of generation of meta-data in both the production and consumption and crucially explores how automation of such generation can enable a more feasible means of constructing these Object-based experiences.

1.5. A New Dawn for Media

As consumer habits trend towards more easily consumable forms of media, traditional broadcast media needs to stay relevant. YouTube[5], Netflix and social media are capturing more and more of the available audience spaces. These platforms succeed through providing immediate and adaptable forms of media with more mobile tailored experiences and a distinct focus on brevity.

Utilising Object-based Broadcasting traditional broadcasters can approach this problem on two fronts:

---

[5] YouTube, https://youtube.com

- through creating more complex, immersive and interactive experiences, exploiting smart technologies and smartphones/tablets/smart speakers to immerse the user
- through building more crafted and tailored experiences for users, with media that can adapt to suit the user's length or screen size requirements.

Object-based Broadcasting has the potential to provide a new dawn for broadcasters and content producers more generally. If I can exploit this ever-increasing interconnectivity of devices, can I create a new form of media which is more informative, compelling and enjoyable for the user?

Historically broadcasters have focussed on improvements in broadcast quality as this was the most pressing challenge for them to work on. This was the obvious focus with the technology available in the home at the time. In the modern home with its screen and sensory overload what should broadcasters focus on? The real challenge in relation to realising Object-based Broadcasting as a concept is in providing adequate inferences and context to a media playback engine. This requires work at both ends of the production pipeline.

In production, I need to generate well-described and functionally descript media assets in a reasonable manner, ensuring that as much contextual information from each captured piece of film and audio is maintained. To do so requires an understanding of:

- what is in each scene
- what sections of the scene are imperative
- which sections provide detail
- which sections provide context
- which scene requires content from another scene to make sense

Providing this information could enable a machine to understand the narrative and enable content makers and story writers to maintain their creative influence over the media while providing a more tailored experience.

In consumption Object-based Broadcasting needs to tame this unbounded and disparate space of smart home technologies. The aforementioned piecemeal nature of the home has given rise to a range of devices and output mediums from a range of manufacturers to be present within the average household. Broadcasters have typically addressed this problem through standards; MPEG, HLG, DASH are all examples of how video and audio have been standardised such that the playback device is largely irrelevant for media consumption.

This raises several questions:

- what standards are required with Object-based Broadcasting consumption?
- how can the disparate nature of the home and its devices be contained in an experience whilst also maintaining their full interactive input/output potential?

As such, we need to develop a set of design requirements for the pipeline or lifecycle of Object-based Media, one which has the potential to satisfy a broad set of the conceptual consumption experiences proferred. I deduce that the following requirements are necessary for production:

- Means of describing and contextualising media assets
- Feasible (both in time and expense) method of generation of meta-data describing assets
- Maintenance of raw assets through to consumption
- Tools to enable creators to craft their experiences

And in Consumption:

- Integration of existing input and output devices
- Abstraction/Standardisation of input and output, making devices more accessible to content creators
- Content delivery network that responds to playback environment and scales to demand

These are the challenges presented to industry, and the barriers to the production of feasible, extensible, and compelling Object-based Media. Industry researchers are already looking at methods of supporting some of these steps, maintaining raw assets with tools to “virtually” edit footage in a non-destructive manner [11].

1.6. Internet of Things

In developing this set of challenges, I wish to highlight the suitability of the Internet of Things as a potential match to these challenges, and as a means to deliver on the promises of Object-based Broadcasting. Its suitability in addressing many of the barriers is through a symmetry in both approach and more explicitly in providing some of the missing data in both production and consumption.

Internet of Things (IoT) as an industrial movement typically refers to connecting previously 'dumb' devices to the Internet, with a view to providing integration of data and services. It has both industrial and consumer applications. Traditional content production processes, it could be argued, have a similar need for integration. Much is tracked by individual stages and roles in the production workflow, however little is maintained or stored in an accessible form (e.g. written down, discarded once finished with). Clearly IoT system architecture and explicit sensing could play a role here.

For example, if we would like to know which scene in a cookery show contains the chef chopping the carrots we need a means to label this information in a cost effective, time efficient and reliable manner. Now some of this information may be captured in scene markers already, though their resolution (start and end time) could be too low for some applications. If smart appliances are already naturalised in the kitchen, can instrumented sensing in production environments similarly provide rich contextual meta-data?

Working from a point of needing to maintain context and be able to infer in some way what is going on within each media asset I look to sensing technologies. The Internet of Things has supported an explosion in domestic and industry sensing technologies, and as such, we can track a vast array of things with small discrete sensors. In productions, I need to label detailed interactions, object-tracking and actions. However, when utilizing ubiquitous computing technologies, are we able to track this information in near real-time, and crucially, can we automate the generation data labelling of these aspects of film?

Realising the challenge that faces broadcasters, I will look at how the Internet of Things can facilitate production of Object-based media and generate detailed labelled assets with minimal manual effort and in keeping with existing production workflows. This sympathetic goal of deeper integration demonstrates a seemingly obvious link to the development of IoT technologies for content production.

The presence of IoT in consumer homes is already well established. Modern homes provide a wealth of sensory information and interaction mediums and look set to continue to provide even more. Broadcasters have a real opportunity in including these devices in their experiences to create more compelling and immersive media for the public to consume. Take two types of experiences that may be improved through integration of IoT into media:

Table 2: Experiences enabled by the Internet of Things.

Experience	Description	Exemplars
Functional Experiences	Developing more useful, intuitive and contextually appropriate experiences.	Weather reports influencing the thermostat controls in the house. kitchen appliances being pre-heated at the right point in a cookery show
Entertainment Experiences	Immersing the user in the experience. Involving more of their environment in the playback. Introducing a more dynamic and compelling experience.	Using the speakers and lighting in the house to provide extra ambience. Moving the content from the main screen onto mobile devices (e.g. smartphones/tablets) could create a whole room/house entertainment. Using movement of the user throughout the house to guide the experience.

The Internet of Things can be viewed as an opportunity for Broadcasters to realise the breadth of opportunities for Object-based Broadcasting.

Broadcasters have already stepped into the fray with Smart Home technologies. For example, the BBC have developed an interactive story experience [20] for Amazon Alexa – using this novel piece of smart home technology in such a way, we can see the industry interest in such experiences. The proliferation of smart home technology is predicated on industry backing. As with any introduction of consumer technology, its success is measured in the content available for it. The iPhone would not have been the success it is today without the App Store. Windows would not have been the success that went before it were it not for mass market business adoption of it and development of a breadth of content for it.

Integration of smart home technologies through platforms such as Apple HomeKit[6] provide a means to enable more comprehensive input/output for end users in their homes. In order to realise some of these more complex immersive playback experiences that move beyond the single/multi-screen setup deeper integration like this is needed.

The Internet of Things offers potential answers to the challenges offered by OBB; its pre-existing presence in the home and the potential for coordinated sensing in productions works in favour of both production and consumption. This natural symmetry at both ends of the production pipeline also potentially lends itself to better integration between the two.

1.7. Data vs Meta-Data

I argue throughout this thesis that data underpins much of these Object-based Broadcasting experiences. Typically data is considered something of direct use as a primary feed of information while meta-data is data that is of tertiary use, describing or providing supplementary information the original data it is related to. In undertaking this research I would refer to video as the primary data source and all other data supporting these experiences as meta-data. This however could be considered a misuse of the term meta-data. Object-based Broadcasting experiences can take on many forms, textual representations, abstract environmental experiences (lighting, sounds) and as well as more typical video supplemented by tertiary data type experiences. At the point of capture the primary data source is unknown, therefore it is inappropriate to refer to

[6] Apple HomeKit, https://apple.com/ios/homekit

any captured data as meta-data. When I discuss collection of this data and I am referring to the creation of data which describes the film assets I will continue to call this meta-data generation, this includes when discussing existing works as this terminology is used by the authors of these works. However when discussing the use of this data to drive OBB experiences I will describe it as data. Sometimes the original source of this data was meta-data but when it is used as part of an OBB experience it is no longer meta-data. I wish to ensure that this work adequately covers a range of OBB experiences and does not prefer any one path in conducting this research.

1.8. Immersion

Content makers strive to make their content immersive drawing consumers into their content and making them yearn for the next episode. Immersion is defined in the Oxford English Dictionary as “deep mental involvement in something”, which can be attained by encouraging attachment to characters and plot lines. Often this is achieved through the application of human psychology, whereby content producers exploit our innate ability to relate and care about other individuals and in some cases more sadistic parts of our psyches. Indeed, often the assessed quality of film which producers intend to replicate is presence. Baños *et al.* [5] reason that presence is a factor of technical immersion and emotive relation to the content. In order to not conflate immersion and presence in this work, I will endeavour to focus on immersion as the desirable quality in Object-based Broadcasting in its generic sense. Improving immersion with OBM will undoubtedly also improve presence in appropriate contexts.

In this research, I made immersion the fundamental quality upon which I focus and improve through Object-based experiences. Immersion is the quality that underpins a significant proportion of historic advances in content broadcast. Increasing picture quality, audio channels and more were in the pursuit of more compelling experiences for user, but I reason this comes through increased immersion. Take for example, a cookery show, here presence is not necessarily the desired quality. Instead, an ability to clearly and adequately follow the demonstrated steps or in some way relate the content to their own environment is more desirable. Similarly, for more informational shows, the bandwidth for absorption of material is perhaps the more desirable quality, arguably these fall within the bounded definition of immersion. Object-based Broadcasting provides a new level of potential immersion in this regard, increasing the availability of input/output mediums, with contextually aware content playback and generally more adaptive media could provide more immersion in our somewhat expanded definition.

1.9. Research Question

In introducing this work, I discuss the focus of much of the broadcast industry upon more tailored content experiences, exploiting the vast array of mediums which are proliferating in the modern home and a constant desire from the consumer for more compelling media experiences. Object-based Broadcasting as a concept provides exciting possibilities for new responsive, reconfigurable and immersive content experiences but is currently without feasible mechanisms to generate the media for broadcast. Distinctly lacking are reliable and cost-effective ways of generating contextual data describing media assets, and a method of delivery and coordination of playback in the home. The Internet of Things looks to provide greater connectivity to existing devices and sensors, providing new or more accessible control and inferences for an environment. Furthermore, smart home technologies introduce a wealth of new input and output devices in the home. In this thesis I will investigate this potential synergy in focus between immersive playback experiences and Smart Home (IoT more broadly) technologies. I will focus on the opportunities for deep infrastructural sensing, input/output and control when working at the nexus of these two industries. In doing so, I hope to investigate the melding of software and media. The following research question guides the research in this thesis:

How can the Internet of Things and Smart Home Technology support the production and consumption of Object-based Media?

Support is the operative word here and refers to supporting the production and consumption of OBM. Achieved through the deployment of ubiquitous computing technologies, with a view to providing a feasible means to construct OBB experiences: making OBM cheaper to produce, both in terms of price and time with deep production sensing and integration, and through providing more compelling content experiences through exploiting the surge in Smart Home technology.

In exploring these factors, this work also raises, and indeed seeks to address, several further questions. Does the Internet of Things help drive this industry-wide shift towards reconfigurable and responsive media with Object-based Broadcasting? Does

IoT address the crucial question of the feasibility of Object-based Broadcasting as a production and consumption concept? This is one of the fundamental roadblocks in the further adoption of such experiences. Focussing solely on the production of media assets or conversely on the consumption would lead to a fundamentally incomplete and undeliverable solution. Only through consideration of an end-to-end solution can I first identify ubiquitous computing's contribution to addressing some of the challenges related to the adoption of OBM and then begin to develop technical solutions to these challenges.

1.10. Objectives

I will now discuss the practical objectives of this research project, seeking to operationalise the research question of this thesis. There are distinct areas of the production pipeline that I target with carefully crafted research deployments. I have outlined these in the section below and throughout this work. Each of these objectives target a specific part of the production pipeline or media lifecycle. In approaching these challenges to the industry with Object-based Broadcasting, I hope to address the feasibility of IoT technologies within broadcast contexts and investigate the nexus of these two cutting-edge industry movements: Object-based Broadcasting, and the Internet of Things. Through my research I seek to provide a comprehensive analysis of the feasibility for ubiquitous sensing and IoT technology, and in doing so, provide an adequate answer for my research question above. The objectives of this research as therefore as follows:

- Run an Internet of Things sensor instrumented production, to establish suitability of the Internet of Things in media asset labelling.
- Explore greater integration of existing production, personal, and infrastructural data to provide greater understanding to generated film.
- Exploit Smart Home and Internet of Things devices in the home to create immersive content experiences.

1.10.1. Objective One: Run an Internet of Things sensor instrumented production, to establish suitability of the Internet of Things in media asset labelling

My first objective looks to tackle the distinct lack of integrated and reliable data describing media assets generated in production. I wish to explore a feasible and reliable means to generate information that contextualised captured media assets. Describing each asset with time-coded data, detailing actions being performed, and other specific information as necessary such as items or ingredients in use in a cookery show. Sensing hardware such as wireless accelerometers and RFID transponders are priced at a level that they can be used to reliably monitor the usage of objects or capture the detail of an act's performance.

In order to realise the opportunities of Object-based Broadcasting, we need greater understanding from media assets. This challenge, I believe, can be addressed through exploiting ubiquitous sensing technologies supported by the infrastructural technology developed for the Internet of Things. In answering this question, I conducted a large deployment with the BBC in a professional production with a vast array of embedded sensing hardware.

1.10.2. Objective Two: Explore greater integration of existing production, personal and infrastructural data to provide greater understanding to generated film

Consumers live in a modern world of measured excellence, we track an enormous amount of information about ourselves for health and convenience features. Furthermore, companies capture a wealth of information about their users, governments collect data about their citizens, councils collect footage and data about their local area. Continuous improvements in smartphone camera technology has led to greater ubiquity in capture of moments for personal reflection. I believe that there is an opportunity for greater understanding from these datasets, both in terms of visual and non-visual datasets. Integration – the process of combining datasets to gain new meaning – could provide compelling contextual information to aid the generation of well-labelled media in both consumer and professional contexts, through open access policies for data sharing and preservation of production data respectively.

Professional broadcasters have been driving this initiative with their R&D departments and lack of major budgetary constraints. However, I should also address the question of supporting OBM content generation for individuals and small production studios. Content production is in the midst of a paradigm shift with smaller production studios seeing larger traction with their content, primarily, it would appear this is due to services such as YouTube RED and Patreon. Through this research objective I will look at the technical challenges associated with small scale OBM production and look at novel media production techniques. Crucially, I wish to explore this more low-tech setup exploiting the wealth of pre-existing information we capture about ourselves and introducing novel technologies to provoke greater data capture. Through increased integration of these existing datasets I see a potentially untapped resource for generating labelled media assets.

1.10.3. Objective Three: Exploit Smart Home and Internet of Things devices in the home to create immersive content experiences

The concept of Object-based Broadcasting encapsulates the idea of an experience being ultimately suited to its environment. In a puristic sense, this manifests itself as making the very best usage of the available input and output mediums an environment has to offer. Input and output in mediums in traditional broadcast media would be TV remotes, perhaps touch screens for input and screens and speakers for output. However, I view an opportunity to provide even greater responsiveness for consumption environments – exploiting Smart Home technology and connected devices with characteristics for output and sensing for input. In doing so I aim to construct more holistic consumption experiences which provides greater suitability to an environment, alongside control interfaces that are evolving to match the increased complexity and degrees of freedom these Object-based experiences offer.

In development of new immersive consumption experiences, I must also look at developing new control interfaces to enable users to adequately control their consumption experience. I will again look at the nexus of these two industries, applying novel control paradigms and mediums to media control and playback respectively. Through researching this end-user space (which may have production application and implications), I aim to complete the end-to-end study of ubiquitous sensing technologies in media productions, providing a comprehensive answer to my research question.

1.11. Methodology

The scope of this thesis is in understanding the contribution that ubiquitous computing technologies can make to the production and consumption of Object-based Broadcasting content. Given the nature of Object-based as a broad and far-reaching concept, this scope is therefore inherently broad. In conducting this research, I have developed a clear methodology to ensure that my work remained focussed, whilst also aiming to avoid arbitrarily restricting the concept of Object-based Broadcasting. I aimed to evaluate IoT technology in a broad range of scenarios, producing lightweight evaluations at each stage.

In each case technologies were deployed in a prototype form; their levels of completeness range from near production quality to entirely experimental one-use bits of equipment and software. Object-based Broadcasting is still in its early evolution, the scope of the concept itself is constantly changing and broadcasters' approaches (certainly the BBC's) are adapting to suit. Content delivery networks, sensing architecture, consumption mediums are all in a constant state of flux; as such, deep evaluations of any one of these prototype solutions would be inappropriate.

1.12. Production Pipeline

In outlining my approach to this research in the Methodology section, I posit that, in order to assess the potential contribution of IoT technologies to OBM, I must investigate its application at each stage of the production pipeline. To more clearly elucidate this methodology and my approach to this problem, I will highlight where each chapter aims to contribute understanding of the melding of IoT and media. To begin with let's look at the pipeline from a traditional media production.

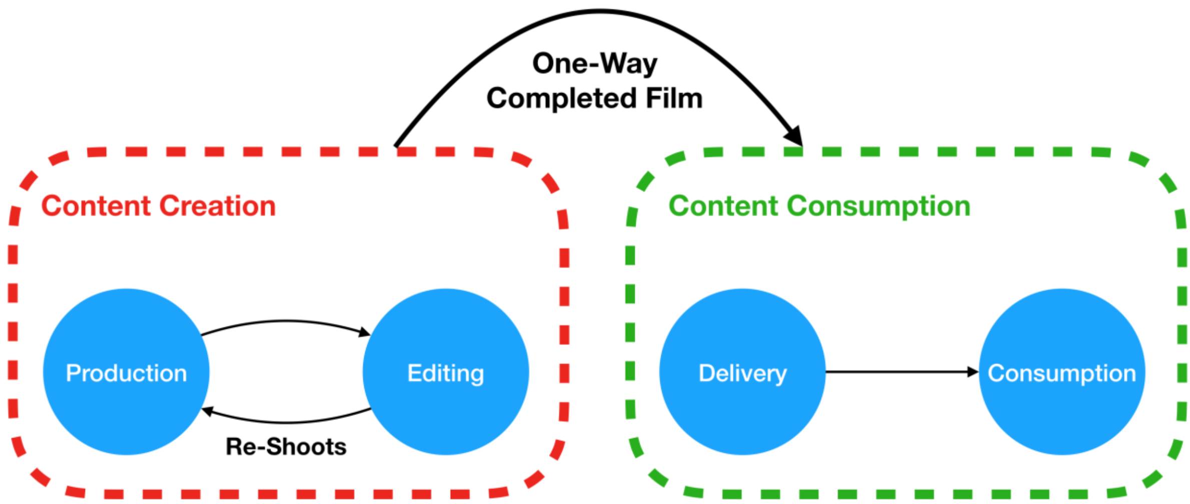


Figure 1: Traditional Media Production Pipeline/Lifecycle.

Figure 1 is a very simplistic view of a production, the stages involved with each of the blue circles in the figure are much more complex and often influence each other. A production pipeline diagram provides an overview of the stages of a media production from project initiation to a user watching the content on their TV or smartphone. Broadly this can be divided into two main stages; Content Creation and Content Consumption. Content Creation stages in a traditional media production by and large do not consider the output medium, beyond it being presented in 16:9 with perhaps 5.1 surround sound and necessary meta-data (info, subtitles, audio description etc.). Similarly, the delivery of content is largely unaffected by the type of content. There is inherent genericism which has evolved over time as broadcasters have determined best practices in relation to content production and broadcast. Similarly, industry needs time to establish best practices in a generic sense for OBM productions, though with adequate extensibility to reach the broad set of consumption possibilities.

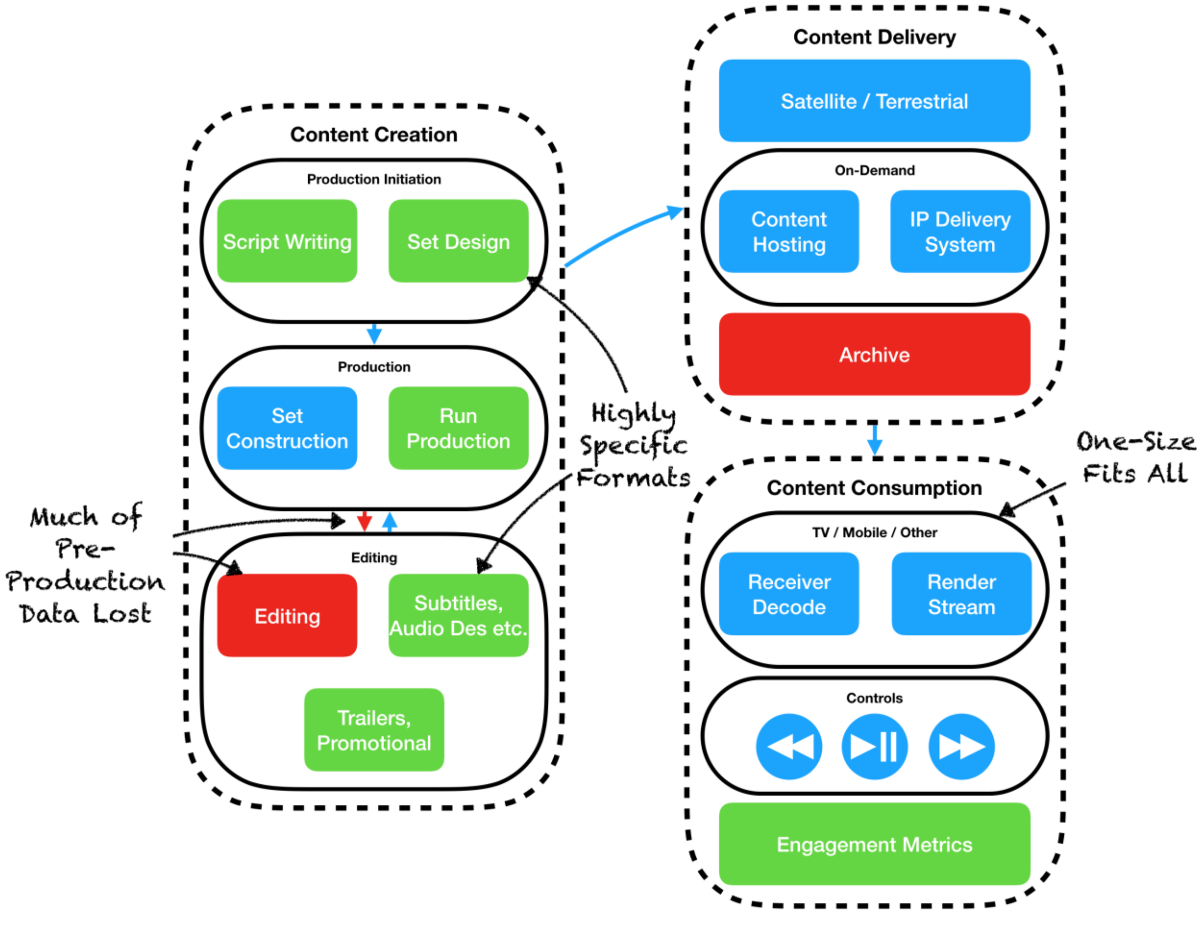


Figure 2: Comprehensive Overview of Traditional Media Production.

Figure 2 shows a far more comprehensive overview of a traditional media production. Each of the sections of the production, shown in Figure 1, have been expanded upon and colour coded. Green items and arrows are sections where data is generated, red where data is lost, and blue are neutral. Throughout this thesis I will look to target each part of the production pipeline, assessing the contribution of Ubiquitous Computing technologies at each stage. At a high level: in each stage highlighted red, I will look to reduce the irreversible destruction of information; where there is green, I will look to amplify the contextual information generated; and where there is blue, I will investigate, where necessary, we can gain further insight. In each case I discuss this as generating further insight with the aim of: constructing compelling Object-based Broadcasting experiences, through deeper Machine Understanding of the environment; and enabling editors to be more creative in their adaptations of content and narrative engines driving experiences.

Any production pipeline I describe relating to an OBB pipeline will be inherently specific to the type of production. The concept of OBB is broad and accommodates a vast swathe of novel consumption experiences. Therefore it would be inappropriate to say that any one pipeline covers all possibilities. However, to guide my research I will suggest that there are some aspects upon which we can focus.

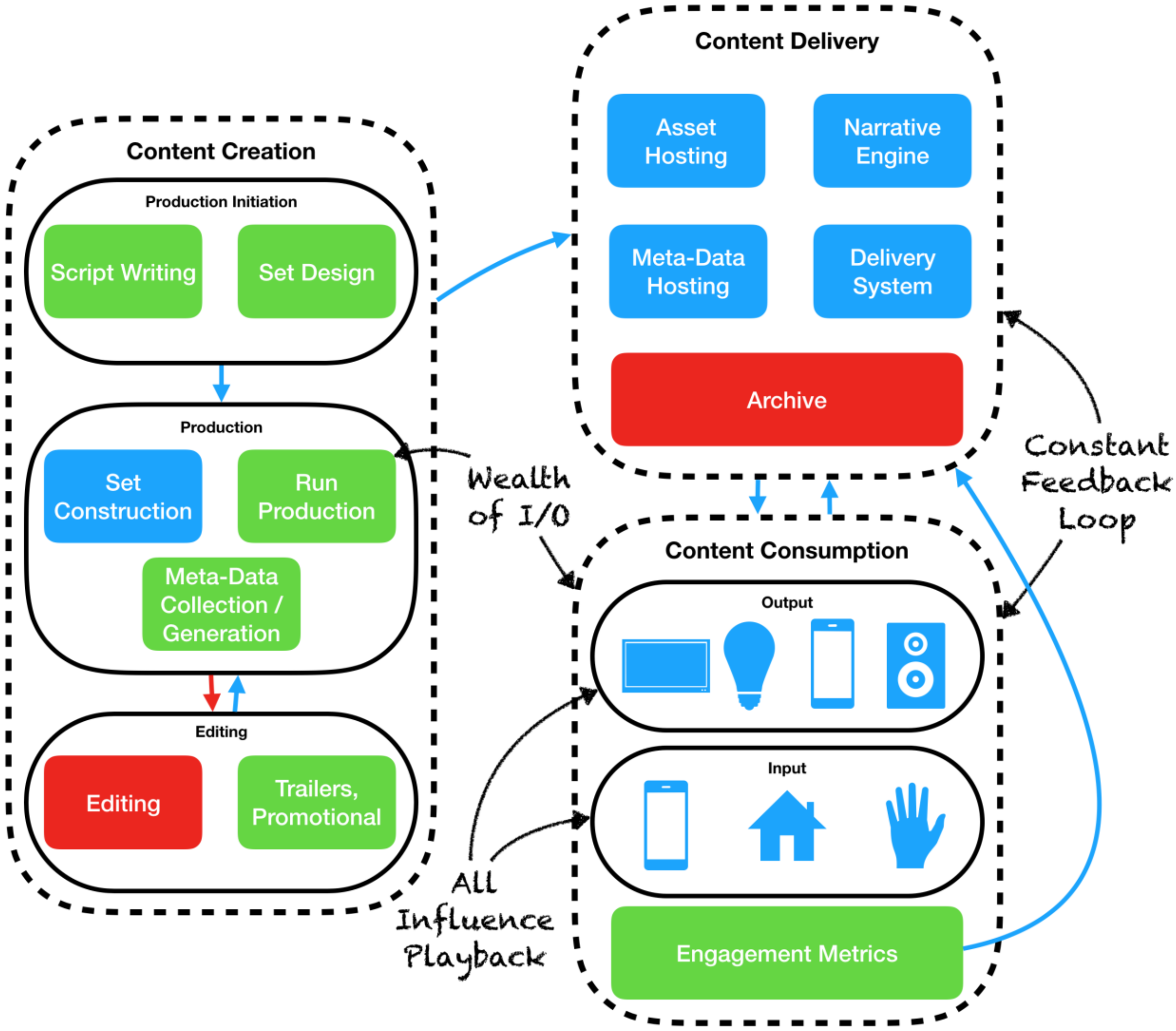


Figure 3: An Exemplar OBB Production Pipeline.

Figure 3 shows an exemplar OBB production pipeline. This layout may not cover all possibilities for Object-based media, however, it provides a good set of foundations for focusing my research. There are some assumptions made here about how data is generated ahead of time, or about the destructiveness of some processes etc. However, I sought to highlight the problem areas in each area with this diagram. Data collection and generation in productions is currently limited and often not automated. If we agree that data describing media assets is one method through which we can realise Object-based experiences, then we need to automate or at least facilitate this process of generation to make it feasible. Processes like editing are inherently destructive; each final editorial decision could limit future application scenarios for Object-based Media. So, should such decisions be reversible? There are so many opportunities to explore in content consumption, and a wealth of devices, and input/output mediums to play with. I will look to explore the challenges surrounding some exemplar experiences. In each relevant subsequent chapter, I have used this model to guide the reader. I have highlighted which sections of the pipeline I am focussing on in each section.

1.13. Thesis Structure

In this thesis, I explored the opportunities for Object-based Broadcasting for the professional broadcasting industry and the consumer. I investigated how sensing technologies can be introduced to the production workflow to facilitate the production and consumption of such media. In exploring these concepts, I worked to exploit the rapid development of smart home and Internet of Things technologies to support these developments in Object-based Broadcasting. Through identifying how existing efforts to support the production of Object-based Media have failed to realise this potential I identify a distinct lack of adequate data describing media assets. I reason that such a lack of information further compounds the development of Object-based Broadcasting as a concept, limiting its wider adoption and deployment.

Through exploiting ubiquitous computing technologies and concepts of woven and embedded hardware and sensing in environments, I look to explore the opportunities of Object-based Broadcasting with sensor driven meta-data generation and application.

I present three case studies which seek to pervade each part of the media production workflow from content creation to consumption. In conducting these studies, I demonstrate through a deploy-and-evaluate-based approach the feasibility of sensing as a means to construct compelling, immersive content experiences that are adaptive and reconfigurable.

The contexts in which I chose to run my research cover a broad range of the media lifecycle. In conducting my research in this way, I designed solutions, gained experience and implemented workflows which address some of the challenges related to the adoption of Object-based Media in production and consumption in professional and non-professional contexts. The concept of Object-based Media is broad, covering a vast array of consumption mediums and scenarios; this universality of the media is one of the core strengths of Object-based Broadcasting as a concept. As such, exploring one siloed application of Object-based Media could mean the results could be considered inconsequential to the development of Object-based Broadcasting. As such, I chose to explore radically different consumption scenarios not only to test the applicability of IoT sensing in a range of contexts for OBM, but also to ensure that this work stayed true to the concept of Object-based Broadcasting as a wide reaching and pervasive industry shift.

1.14. Summary

This introductory chapter has provided some insight into the concept of Object-based Broadcasting. It has detailed my research questions as they pertain to this research. Each research question is clearly addressed in this thesis and relevant chapters will highlight this. I detailed the rationale and goals of the proliferation of this exciting reconfigurable and responsive media.

This thesis aims to answer the question of whether ubiquitous computing technologies have a place in media productions and consumption with a view to supporting the production of novel, immersive and responsive media experiences. I highlight my motivation as that of a lack adequate data describing media assets, destructive media production processes and with a focus on integration of existing and new datasets as an approach to solving this fundamental barrier to more adaptive and reconfigurable media experiences. In doing so, I focus my work on the issue of reliable and feasible means of generating labelled media assets, investigating deeper integration of existing sensor data and development of adequate control interfaces for these novel experiences.

Related Work

Chapter Two

Open Lab at Newcastle University

BBC

Chapter 2. Related Work

In this chapter, I looked to emphasise the motivation behind this thesis. I investigated the appropriateness of the Internet of Things, and ubiquitous computing technologies more generally, to support Object-based Broadcasting as a concept. I begin by looking at the foundations of Object-based Broadcasting, its founding literature, and conduct an in-depth review of existing approaches to like-minded concepts and meta-data generation and collection in film. I also cover existing adaptive media experiences, looking at their means of deployment, the control available and narrative management. Bringing in the subject of the Internet of Things and Smart Home technology as a new playground for broadcasters to create immersive experiences, exploiting the ubiquity of input and output such technologies offer. I discuss the potential for development of new control techniques to offer more nuanced and detailed control to this shift in media experiences.

In doing so, I hope to elucidate the notion of advances in reconfigurable and immersive media being inextricably linked to development of data/meta-data rich raw media assets. I see Object-based Broadcasting as a conceptualised response to a consumer desire for more responsive and reconfigurable media experiences and broadcasters providing a practical concept to work from.

2.1. Object-based Broadcasting

Object-based Broadcasting (OBB) [19,84] is a response to new consumer content demands on the production of broadcast media. Traditional broadcast pipelines, in both radio and video, broadcast a packaged, edited, single stream of information to all users regardless of playback device or environmental factors.

OBB looks to rethink this workflow, offering media as a set of assets with associated data describing each asset. This data is defined as supplementary information that provides description of the scene, camera information, location information or contextual information about what is happening in each video shot.

In the most generic case, OBB assumes no pre-determined use case for any media. This enables elements of the consumption experience (such as timing, narrative and display size) to be decided not by the broadcaster during post-production, but by the consumer, their device or environment (or a combination of these) based on narrative rules or external parameters.

As an example, consider responsive web design – developers build their sites to resize, re-order and reconfigure based upon the available screen real-estate. A website will look different on a desktop compared to a mobile device due to a set of rules put in place by the developer. The developer did not need to redesign the website for every single display size when creating the content. OBB can be thought of in a similar manner, where a set of rules or protocols defines how the media objects should be combined, allowing us to achieve a similar level of responsiveness by re-framing the media to fit the device. This could be as simple as presenting a traditional wide-screen shot to desktops but an appropriately reframed portrait feed could be shown on a mobile device.

Some more complex examples may include:

- redistributing the content across an array of displays in an environment or other environmental devices such as lighting
- adapting to individual user needs with character re-introductions in TV shows
- adapting a cookery show to accommodate multiple consumers cooking along together

During a TV show, custom catch-up experiences could be presented to users combining salient sections of previous shows into the latest episode to allow users to more easily catch-up or understand the context behind scenes they missed, and this could be personalised to each user based upon their viewing history.

In essence, Object-based Broadcasting is moving the onus of presentation from a carefully constructed output from an editing and production team to a set of well described assets. Each asset has a place, as before, but is labelled in such a way to describe how it can be used, where its place in the narrative is, what elements are in the scene and their significance. Such an approach enables a more responsive approach at the playback 'edge' (the users TV set, smartphone or combination in a living room). The experience can be adapted in response to the presence of certain input or output mediums with a logic model deciding how the individual assets are finally presented.

To achieve such a result, careful labelling of each asset (video feed, audio channel, pictures, text) is required, this information needs to be as detailed as possible to enable the device or user to make informed decisions about what to display. Each camera shot needs a careful description of the view, actions in the scene and potentially device (e.g. appliance) usage.

Previous work on OBB has approached this issue through manual labelling of media. Churnside *et al.* [18] experimented with reconfigurable radio shows; sections of elaboration could be varied depending on the available time for the user. This content was produced from an existing radio broadcast. Logical dependencies from parts of the story were produced and sections labelled as "introduces", "expands" or "resolves". A narrative graph was then produced to encapsulate the production. Cox *et al.* [22] looked to produce some bespoke OBB content. Their system, "CAKE", was an interactive cooking show which allowed viewers to select individual parts of a recipe to cook and swap some ingredients for others. The viewer experience reconfigures in response to these selections. Again, the data/meta-data to support this experience was constructed manually as a post-production process. This not only restricted the possibility for configuration as each scenario has to be known beforehand but also requires a significant amount of manual labelling of video in post-production. They successfully demonstrated an Object-based production; however, they lacked a sustainable means to construct the rich content required for the consumption scenario.

2.2. Producing Descriptive Data

The specificity of each consumption scenario is a recognised constraint of current OBB workflows, and recent work as part of the 2Immerse[7] [60] project has approached how production tooling and workflows must adapt to new forms of media production to allow for a sustainable workflow. The requirement to generate rich context based data for use in these reconfigurable contexts is an ongoing research field, roughly split into two areas: post-production generation; and point-of-capture generation.

---

[7] 2-IMMERSE, https://2immerse.eu/

2.2.1. Post Production

Previous work in automated labelling and meta-data gathering in video has focussed on Computer Vision. Object detection within regions of an image was originally explored by Duygulu *et al.* [27] and Barnard *et al.* [6], where the authors could identify distinct objects and demonstrated reasonable accuracy in generic objects (planes, animals, scenery). This was expanded to full scene descriptions by Kulkarni *et al.* [52], he demonstrated a means to construct full scene descriptions combining previous work to describe objects within the scene and using a Constructed Conditional Random Field to produce full scene descriptions. This work was refined by Mitchell *et al.* [61] to reduce the often "noisy" descriptions produced by Kulkarni's method and similar processes have been successfully commercialised. Google's Vision API[8] and Microsoft's Cognitive Services[9] both offer highly accurate image descriptions trained on their extensive image search meta-data. Open Source Machine Intelligence project TensorFlow[10] demonstrates continuing improvements in image descriptions and performance [83]. Hu *et al.* [43] then expanded this work to video, using generated descriptions to automate the detection of unusual activity on CCTV cameras and real-time video scene analysis is also available through Microsoft's Cognitive Services. Although these techniques are now becoming highly accurate, they lack detailed information on non-visible context such as performative action and intent, which can only be obtained during the capture process.

This post-production approach to meta-data re-generation, as I will describe it, can be considered the current gold standard. Indeed, Amazon has the most comprehensive example of this with its X-Ray[11] for Prime Video feature. X-Ray provides time-sequenced meta-data to a large catalogue of the broadcasters shows – akin to some screen experiences I describe below – cast information, story embellishments and supplementary videos are provided. According to Wired [77], "computers can do some of the identification and time-coding, but every movie or TV show has some human touch as well", vision techniques can provide some of the ground work but fails to understand the nuances to deliver experiences such as this in totality. The authors of the article emphasise the issue with "that's why there are only about 100 titles a week

[8] Google Cloud Vision API, https://cloud.google.com/vision/
[9] Microsoft Cognitive Services, https://azure.microsoft.com/en-us/services/cognitive-services/
[10] TensorFlow – Machine Intelligence Project, https://www.tensorflow.org
[11] Amazon Prime Video X-Ray, https://www.amazon.com/gp/adlp/xrayvideo?tab=Trivia

added to X-Ray". This quote succinctly surmises the problems with Object-based Broadcasting experiences, manual labelling will always lead to the content being limited.

2.2.2. Point-of-Capture

Media productions already mandate a range of procedures and roles to ensure the smooth running of a production. For example, continuity, the principle of artifacts, context and narrative soundness being maintained throughout a production typically requires a continuity editor on set. This role along with the likes of shot supervisors are often setup as very manual roles, even paper based. Furthermore, production crews will often capture segments from multiple angles and with subtle variations on the content to 'cover all the angles' at the shoot. This example of maximising captured detail for later filtering is emblematic of how Object-based productions should be run.

Object identification at point of capture has been well researched beyond vision. Red-tag [7] takes a different approach to object identification. The authors look to detect which objects are currently within a scene using IR emitters flashing in an encoded pattern mounted on objects and people in the scene. This enables timestamp identification of labelled objects recorded on any camera pointed into the scene. In an alternative approach, Bartindale *et al.* [8] demonstrate that pre-defining some context based meta-data before capturing video by using on-screen templates of shots on a mobile device can produce reliable context based meta-data, but this approach is limited in the type of context that can be recorded (i.e. shot angle, label of who is in shot).

Automatic meta-data generation systems are widely used in sports broadcasting. Football statistics (Opta Stats[12]) and goal-line technology (Hawk-Eye Goal Line Technology[13]) supply a wealth of real-time data that provides a compelling consumption experience for viewers and real-time assistance to referees in matches. In film production, applications to support script supervisors also demonstrate the opportunities for digitised meta-data capture, albeit human powered. Tablet applications such as LockitScript[14] and Script Evolution[15] allow the digital recording of

---

[12] Opta Premier League, http://www.optasports.com/events/premier-league.aspx
[13] Hawk-Eye Goal Line Technology, https://www.hawkeyeinnovations.com/products/ball-tracking/goal-line-technology
[14] LockItScript, https://lockitnetwork.com/apps/
[15] Script Evolution, http://www.simpleapp.fr/evolution/

script notes and time cards at the point of capture. Such tools however are role specific and do not form part of a coordinated output of production meta-data used outside of the shoot.

Current literature thus demonstrates advances in object detection and recognition, but identifying state and interactions being performed is out of reach. For example, in a cookery show, the speed of a blender, the amount of water used, the current utensil in use are all valuable pieces of information that are required to create rich consumption experiences, which are difficult to obtain with Computer Vision.

2.3. Leveraging Context Based Data in Production

Examples of using data to drive a production are also prominent in sports broadcasting. These solutions mostly relate to identifying salient parts of a stream such as goals, in order to generate automatic highlights [3,25]. Using specific features (goal posts, crowd movement etc) they can determine the most interesting parts of a game and automatically summarise them in a variety of formats.

Indeed much of the current literature focusses on attempting to understand where the most interesting segments of the video are; work such as Virtual Director [50] have attempted to fully automate the editing process by framing close-up shots from wide-angle high resolution camera images based on contextual meta-data.

Similarly, Schofield *et al.'s* [81] 'Bootlegger' smartphone application coordinates a 'crew' of users covering a live event, assisting them in producing a quality film with shot overlay templates and other shot guidance. Recruited members choose from a set of roles defined by the shoot organiser and an 'auto-director' assigns shots to be taken to appropriate roles, in a live filming scenario the 'auto-director' ensures complete coverage of the event and provides a complete set of shots (from start to finish of the production) to be composed into a film later based on the context based meta-data captured by the devices about what is being shot in real-time.

Rather than human powered, Kaiser *et al.* have investigated using Computer Vision to make shot decisions by framing shots out of a wide-angle camera placed within a soccer stadium [48]. They attempted to track the position of players, the ball and rapid movement, creating shot decisions based on these parameters. However, they found after much experimentation that Computer Vision alone struggles to provide sufficient information to make informed shot decisions. In reality, the resulting media was mostly of the same thing, and did not relate directly to the action. Kaiser *et al.* [49] also explored how the narrative of a production should be altered when using 360° high-resolution cameras throughout a theatre production's stage. A performance was conducted to explore this new perspective on theatre productions, essentially moving the audience to the middle of the stage. The authors deployed similar Computer Vision techniques to attempt to automate shot selection and guide the viewers' attention. Furthermore, there has been some recent consumer products like Mevo[16], a 4K wide-angle camera with live shot framing for livestreaming services like Facebook Live[17]. BBC Primer [11], a real-time web-based shot framing system hopes to utilize 4K unmanned cameras set-up to cover a wide shot. Thereby allowing directors to change the shot simply by framing out a new segment. Given all the work in this space however, these scenarios are still limited by the availability of reliable and detailed meta-data to facilitate each production process.

2.4. Leveraging Meta-Data to Drive New Consumption Scenarios

Media with associated meta-data has historically been used to drive a wide range of consumption scenarios, from some of the first usages with 'Interactive TV' services such as 'red button' [34] to newer 'second screen' experiences. A classic example is interactive and branching narrative experiences, they require carefully constructed and labelled media sets to enable a rich and meaningful experience. Ursu *et al.* [91,92] initially explored the possibilities of this genre for interactive TV and produced 'Shape Shifting TV'. The authors explored atomisation of content, enabling the media to be reordered with a lower impact on the narrative and explored reconfigurable media within documentary production [93]. However, these workflows required a large amount of additional pre-production work from that of traditional TV. Zsombori *et al.* [107] continued this work by attempting to leverage user-generated content and narrative building to circumvent such limitations, however the style and production value of such content is limited. 'TryFilm' [9] approached the problem by allowing the cast and crew of an interactive narrative production to interact with the content on

[16] Mevo Livestream Camera, https://getmevo.com
[17] Facebook Live, https://live.fb.com

location. The intention was to bootstrap the post-production process by creating useful data on the shoot, however this proved difficult in practice.

More recently we have seen a concerted effort by broadcasters to deliver detailed and immersive second screen experiences. Indeed, Pablo *et al.* & Obrist *et al.* explore these interactive and second screen experiences [16,17,69] envisaging the second screen as a device that compliments the viewing experience both in terms of providing control and extra information about the current content as well as social media integration.

Many consumers are moving away from traditional forms of media consumption with severe declines in set-top boxes in homes and the consumption of Live TV [88,104], instead preferring more convenient and discoverable media. On-demand services like Netflix[18] and Amazon Prime Instant Video[19] allow for a much more flexible and accessible means to enjoy media. The delivery of these services via the Internet also opens up new levels of bandwidth and control that enable Object-Based Broadcasting [19,84]. The wide availability of consumer devices such as Google Chromecast[20] and Amazon Fire TV[21] also enable a cheap and convenient way to provide On-Demand and potentially interactive services to existing TV consumers thus providing a practical means to create the immersive media experiences made possible by OBB.

2.5. Opportunities for Sensing

Thus far, I have focussed on the wealth of existing work in media and content meta-data, reconfiguration and novel consumption experiences. In doing so, I have highlighted a recurring theme of impracticality in scale, depth of reconfiguration and consumption possibilities. Previous approaches have focussed on manual labelling of assets, working from a point of knowing the desired output. Or working on a context recovery process post-production, scraping from what little information remains of the original meaning from the edited, produced output. Some works have looked at more explicit design for adaptive outputs, RedTag [7] provided object-tracking by instrumenting parts of the production studio. Building upon works such as RedTag there is a distinct opportunity

[18] Netflix, https://netflix.com
[19] Amazon Prime Instant Video, https://amazon.co.uk/video
[20] Google Chromecast, https://store.google.com/product/chromecast
[21] Amazon Fire TV, https://www.amazon.co.uk/Fire-Tv-With-4K-Ultra-HD-And-Alexa-Voice-Remote-Streaming-Media-Player/dp/B06XTWLSRF

for sensing, in terms of supporting the design for the unknown. In its most simplistic sense this would involve looking to maintain as much information as possible from the production, avoiding the need for recovery. Thus maximising the possibilities for deployment – in essence capture everything to enable as many uses as possible. Working from this point, how can I design productions in such a way that this is practical and feasible? We are already seeing a fundamental shift in the breadth of technology (input and ouput) in our lives with instrumentation with sensing in the home and industry with the Internet of Things – it is suddenly cheap and feasible to explicitly instrument devices and objects. What role can sensing play in addressing some of these problems in production and how can this assist in novel consumption experiences?

2.5.1. Sensing in the Environment & Internet of Things

In a visual media production, the primary sensor is obviously the cameras, at least in traditional productions. Inherently, vision based descriptive data generation solutions lack the ability to describe actions and processes that are not visually captured in the media. The emergence of IoT an opportunity to achieve ubiquitous sensing in environments, it is therefore within reach to envision consumption scenarios which make use of IoT devices as part of the viewing experience. When considering the production studio as an embedded sensing environment, the viability of IoT for capture as well as output is immediately made clear. Current applications of IoT have largely been confined to two main areas:

- Smart-Homes: Connecting hardware within the home to the internet enabling new levels of control and automation of devices such as thermostats (Nest Thermostat[22]), lighting (Philips Hue[23]), locks (August Smart Lock[24]) and appliances (Smarter Kettle[25]).

[22] Nest Learning Thermostat, https://nest.com/thermostat
[23] Philips Hue, https://www2.meethue.com
[24] August Smart Lock, https://august.com
[25] Smarter Kettle, https://smarter.am

- Industrial Sensing: Deploying cheap, reliable sensing equipment for real-time reporting and analysis of production lines[26], manufacturing infrastructure [24] and asset monitoring[27].

The latter application, while lesser known, is far more prevalent. Previously manual processes such as meter reading, and parcel tracking, for example can now be automated and validated by embedded technologies in the environment. Such sensors could be used to provide the detailed data required to achieve Object Based Media. Indeed it is at this point that much of the meta-data we have discussed can become the primary data source for consumption, becoming data no longer just meta-data. Since many filming scenarios are based around the same constraints as consumer IoT (i.e. the home), the approach of smart environments provides a natural symmetry to smart production environments. There is a significant investment from major technologies in smart technology for the home that can provide assistance to this approach to smart productions. Also, in a practical sense, this symmetry grounds reconfigurable media in what is possible for consumers while at the same time potentially reducing production costs through taking advantage of available consumer technology.

In recent work, Churnside *et al.* and Cox *et al.* [18,22] discuss the possibility of IoT for consuming responsive media, where the playback system responds to actions from the consumer to adjust what is played. Whilst not producing media, the Ambient Kitchen by Olivier *et al.* explored using a sensor instrumented kitchen to gain context from user actions in a kitchen and aid patients with dementia doing tasks [41]. This project drew inspiration from Ambient Intelligence proposed by Tscheligi *et al.* [90], a new concept of user centred design. They theorised environemnts should be sensitive to a users need and adapting or providing feedback accordingly. Olivier *et al.* and Chong & Olivier [72,74] were able to perform stage detection using the sensors within the environment helping dementia patients to complete tasks. They did however struggle with detecting true task completion and action qualities and the work required excessive instrumentation of the environment to achieve this. Further work based in the Ambient Kitchen investigated the application of such technology to assist in language learning

[26] Microsoft KUKA, https://www.microsoft.com/en-us/cloud-platform/customer-stories-kuka-robotics
[27] BlackBerry IoT – Asset Tracking, http://uk.blackberry.com/internet-of-things/platform/application-modules/vertical-applications/asset-tracking.html

[42,78,82] through cooking. The system cued participants with ingredients or tasks in a new language which had to be identified/used before the recipe continued. These works demonstrate the power offered by connected environments with many of the application scenarios immediately applicable to media consumption in a connected environment. Instrumentation of production studios and capturing using IoT technologies provides a cheap readily available source of sensor data to exploit existing production data sources (camera positioning, props, appliances etc.) as well as new potential data sources (positional information and interaction data) to describe the created media.

2.5.2. A Synergy of Focus

In developing IoT hardware, researchers and manufacturers have typically looked to address an issue of a lack of information or lack of information in a timely and contextually appropriate manner. Networked physical sensors and controls enable the flow of timely detailed contextually defined information. Other forms of networked output mechanisms facilitate response to stimuli from sensors. This methodology has enabled automation of many previously manual tasks in industry and created a connected home with greater opportunities for creativity, access to information and control in the home.

Object-based productions, such as those discussed above, and others discussed in this thesis require us to move beyond video as the sole data source. Supplementary data describing other data (video), is what is required here. I posit that the data required to develop some of these experiences goes beyond what is currently possible with the edited output or even the raw video assets themselves. Current vision technologies produce output that is too error-prone, descriptions of scenes that are too noisy and lack detail.

Data that is useful to constructing Object-based experiences are:

- Location: knowing where objects or individuals are within a scene can be useful in establishing who to focus on, what can be framed in or out of a shot etc.
- Usage Information: knowing which items and devices are in use and where the devices are at any one time.
- Explicit Detail: what exactly a person is doing within a scene.

- Scene/Action Significance: the significance of a region (spatial and chronological) of the film to the narrative or understanding.
- Dependency: how a particular scene is related to another. A scene may be an elaboration of another, required to understand a subsequent or previous scene.

This is not a comprehensive list; such a list would be difficult (almost impossible at this stage) to produce as the potential number of output configurations is unknown. However, working from this subset of data we can develop most of the existing examples of Object-based experiences from the literature above and develop novel reconfigurable experiences. Object-based experiences will typically focus upon Scene/Action Significance and Dependency as fundamental characteristics to maintain the narrative of any piece of film. Calculation of this data may be as a function of some of the other data listed above (similar to the process of feature engineering in Machine Learning). Structurally, a sympathetic architecture to support such dynamism in the media would be that of the Internet of Things. Many of these requirements are sympathetic to those offered by IoT and Smart Home infrastructure and devices. There is a synergy that could suggest a similar approach could be used to support Object-based Broadcasting productions. Indeed, many of the sensors and Smart Home hardware being developed for consumer and industry are designed to detect and resolve some of the queries I outline above. In facilitating this kind of information capture for the home how can this similarly be applied to professional productions for data capture and controlling experiences during playback in the home.

The ever-increasing prevalence of IoT devices in the home means that consumers are more familiar with certain devices like Google Homes, smart appliances etc. being present in production sets. Media productions can utilise these in-home devices on set as they are now naturalised. They can be included in scripts as actionable items potentially providing explicit links to symmetric hardware in the playback environment. Sets and productions in many cases are designed around a symmetry with the home, particularly how-to or cookery shows. Exploiting this symmetry to include IoT devices explicitly in set design and the production more generally should provide for an audience receptive to these devices being used on set and generally being present in productions. For example, smartwatches on the wrist of a chef or Google Home being used to set a timer.

This new-found ubiquity and proliferation of IoT and Smart Home devices has led industry to develop infrastructure to support the exponential growth in data needing to be analysed, responded to and stored. Content delivery is a recurring concern with Object-based experiences, delivering an experience to end users when each one may receive slightly different content, could result in a quagmire of issues when it comes to delivering content at scale. Exploiting advances in managing large volume of sensor data will be key to the development of Object-based experiences.

2.6. Summary

Previous approaches to the development of novel consumption experiences and precursors to Object-based experiences have always relied on supplementary time-sequenced data describing the underlying media assets. This data provides the required contextual information to realise these experiences. In each of the works I have documented above, researchers have manually labelled and otherwise instructed the media to fulfil its pre-defined task. Through continual advances in Artificial Intelligence combined with rich and more detailed ingested data to build logic models upon, I believe there is a real opportunity to realise the dream of Object-based Broadcasting as a generic concept without tailoring experiences to specific output mediums. Crucially, I propose, this is through deeper labelling of assets to be ingested by ever complex logic models.

Meta-data for the purposes of content production (auto-directors and assisted editors) have also relied heavily on contextual information to make their decisions. Indeed, many of the production applications of meta-data outlined in this chapter have used a range of techniques to regenerate this content post-production such as: Computer Vision, manual labelling and consumer metrics. Each of the works I discuss above are largely hamstrung by this approach, they are limited in what contextual information they can accurately restore. I argue that maintaining this contextual information through deep infrastructural sensing and integration of services will enable not only these works to realise their potential but open up new possibilities for automation and tooling for content makers.

Simply put, I believe these experiences and production technologies can be reliably and reasonably constructed and supported through increased automated descriptive data generation with ubiquitous computing, IoT and Smart Home technologies.

Immersion is the cornerstone of Object-based experiences. In this work I expand the definition of immersion beyond the emotive and feeling of presence to encompass a broader set of media scenarios. In doing so, I hope to focus on some of the consumption experiences developed on this fundamental quality of media consumption.

In the following chapters I will explore four extensive case and user studies of technologies to support Object-based Media. Each of these studies are purposefully very different. Object-based Broadcasting in its truest sense of a fundamentally reconfigurable playback experience mandates significant advancements in media production technology, content delivery networks and consumer technology. When considering this in conjunction with my stated goals in investigating how the Internet of Things can support Object-based Broadcasting end to end I worked on case studies covering each area.

In the following chapter Media of Things, I explore how IoT sensing can be woven into the production workflow, generating detailed data from a production which could be used to create complex consumption experiences where contextual information about each media asset needs to be known. This builds upon work outlined in this chapter, advancing concepts like RedTag and the Ambient Kitchen providing the meta-data to realise projects such as Kaiser's Auto-Director.

In the chapter Sensed Player I worked on building a compelling Object-based Broadcasting experience that was at, what I considered to be, the most advanced edge of what was possible with the Media of Things dataset. Building upon this rich data set from my first case study I wished to construct a fundamentally reconfigurable playback experience. I sought not only to experiment with content storage and content delivery processes in OBB, but also to gain valuable insight from participants into how compelling these experiences are for consumers.

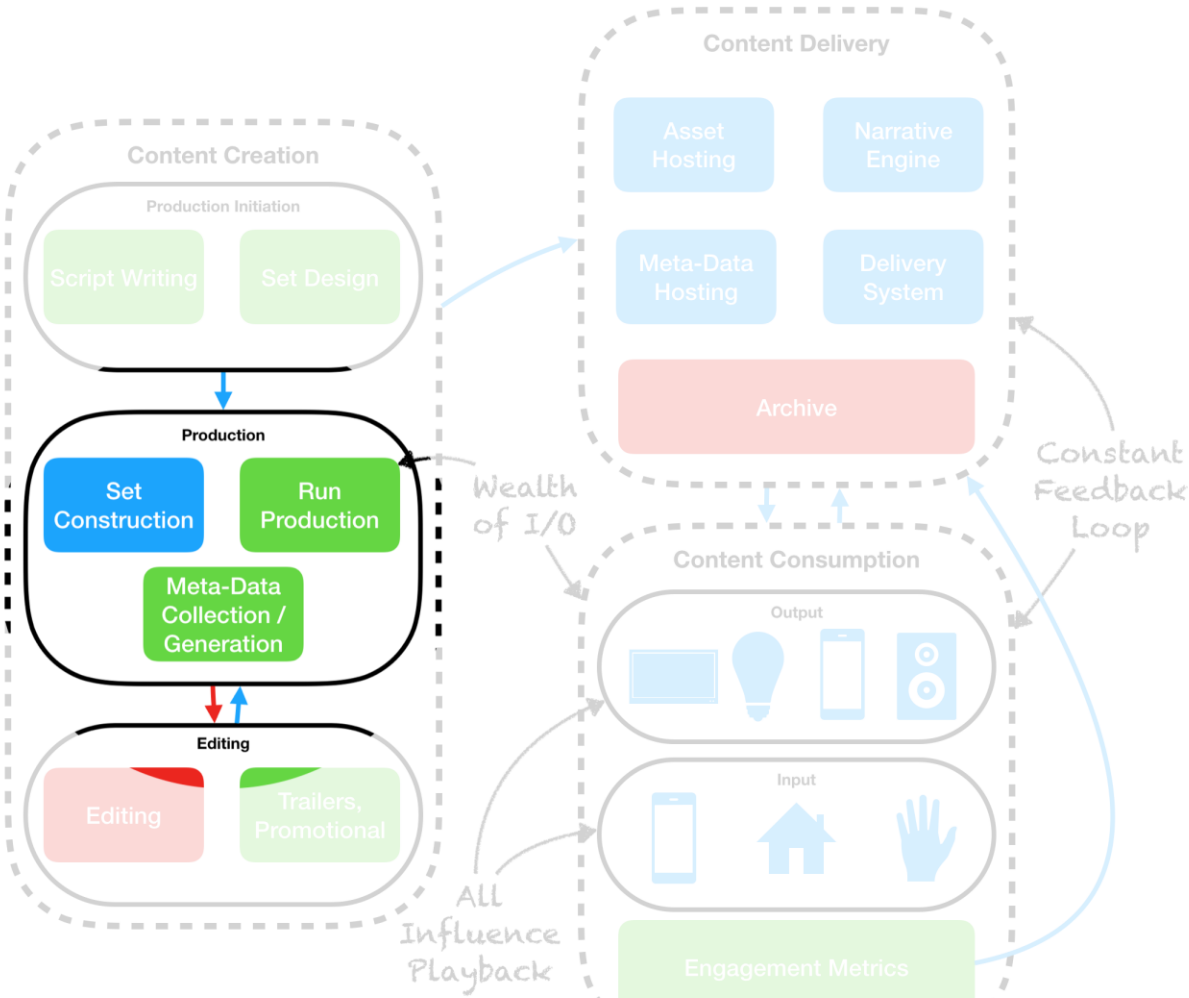
In the chapter Integration, I conducted a case study to develop production processes for consumers to build Object-based Media for themselves. In productions we can build bespoke sensing technologies for each requirement. However in a consumer setting this could be prohibitively expensive to replicate or even come close to parity. As such, I decided to look at how we could exploit existing collected dataset to create Object-based Media. In developing a novel technical solution to enable further sharing of infrastructural data from CCTV cameras with CryptoCam I worked to build a cryptographically secure local sharing technology. This was then applied in conjunction with user generated meta-data to index the media collected from these fixed cameras.

Finally, in culmination of my work, I wished to investigate how in such immersive Object-based Media experiences we can maintain control. In expanding the reconfigurability and number of devices involved in a playback experience we naturally increase the variables under control. This could lead to an overwhelming experience for the consumer, traditional media controls (play/pause/fast-forward/rewind) will no longer suffice. I constructed a Smart Home control system that could enable consumers to regain control over the media.

In conducting these case and user studies I hoped to answer my research questions. Embarking upon a feasibility assessment through a series of case studies of the possibilities for the Internet of Things to support this evolution in media consumption and generation. Production and consumption of media are inherently linked with Object-based Broadcasting, the possibilities for output are limited by what was captured for input. It is only by exploring this research topic from end to end that I can provide appropriate answers to the research questions and objectives.

Media of Things

Chapter Three



Open Lab at Newcastle University

BBC

Chapter 3. Media of Things

In introducing this work, I discuss the potential opportunities for exploiting ubiquitous sensing technologies in productions. The current discrepancy between traditionally produced output media and the descriptive data laden media we wish to produce requires further investigation. Appropriately I should focus on developing media assets with a greater level of detailed data generated at the point of capture. Assessing the contribution of IoT sensors and Smart Home technologies to automated labelling of key details of recorded media assets is warranted. Focussing on instrumenting productions is a natural potential solution – media productions are carefully staged environments. Set and prop designers construct in intricate detail their visions for a production into reality. This staged environment naturally lends itself to a set-up of explicit sensing. Instrumentation of relevant objects is much more feasible especially with prior knowledge of the significance of certain objects. Contextual information is the backbone of these Object-based experiences. Providing content makers with the logical information to construct responsive and reconfigurable experiences. Exploiting the advances of Artificial Intelligence and ubiquitous computing technologies, how can I provide the features for us to make these inferences?

In this chapter I look to tackle my first research objective: can ubiquitous sensing technologies provide adequate contextual information to Object-based experiences? In order to investigate this, I sought to deploy a wealth of sensing technologies within a production studio. I deployed an Internet of Things architecture with disparate sensing, centralised storage, processing and integration. I used a web of devices throughout the production environment with a structured output. Crucially I wanted to go beyond what is currently possible with vision, as highlighted in the related work above. Typically, vision struggled to capture the nuances of certain actions or identify items in a reliable manner. At the conclusion of this chapter I aim to have addressed the sensibility of IoT sensing in production environments.

3.1. Introduction

Broadcast TV technology has evolved at an incredible pace over the past century. Broadcast media is no longer limited to the TV placed in the living room, where the family gathers around a singular broadcast. The availability of content via a myriad devices and platforms has enabled access to content never envisioned by earlier broadcast production pipelines, and as such, the industry is struggling to keep up with the array of available consumption scenarios [70]. Current production workflows are predicated on a singular linear output, usually a TV broadcast, and so the media they produce is often without the necessary descriptive information at the correct granularity to be reconfigured. To support emergent consumption scenarios such as interactive narrative, personalized and multi-device content, new processes and workflows must be developed. Core to these workflows is a rich description of the media being reconfigured to enable reasonable and appropriate reconfiguration decisions to be made.

This move towards treating media as reconfigurable descriptive objects is an approach adopted by many in both industry and academia, and is understood as Object Based Media (OBM) and the associated production processes as Object Based Broadcasting (OBB). Currently, what little OBM that is produced by broadcasters is generated using resource intensive post-production workflows for specific outputs (such as companion apps). By following non-destructive practices to media production and delivery, broadcasters can enable future reconfiguration and manipulation of the media allowing for emergent consumption scenarios. However, without detailed data associated with each media object, reconfiguration opportunities are significantly limited. Although the industry has responded by streamlining end-to-end production through: digitization and standardization of timecode; licensing and the standard inclusion of some semantic meta-data [109], much descriptive information, and in particular - the context in which the source video was shot, is lost. This information is either never captured, captured in a non-machine-readable manner (i.e. clapper boards or hand-written notes), or is lost during the traditionally destructive (i.e. original source media is not kept) production workflow. Combining this data and filling the gaps that exist in context and meaning of production assets is vital in the adoption and proliferation of OBM in the production pipeline.

Drawing from the field of IoT (Internet of Things), I present our open-source, embedded, sensor-based, data capture solution which integrates within a film set. Through discussion of its design, I demonstrate how embedded ubiquitous sensing technologies can be leveraged to produce context rich, time sequenced data at point-of-capture in a studio scenario. I achieve this by reporting on a real-world production where I deployed this system in conjunction with traditional production roles. I discuss how such a system can semi-automate the process of context and more general descriptive data capture, and how such data-capture roles should be introduced as core parts of the production workflow.

3.2. The Role of Data

Reconfigurable media affords new opportunities for creativity in film making and broadcast media. Media is no longer created for one consumption scenario (such as a TV broadcast), in one linear pre-packaged form; instead it is adapted dynamically in response to the context of consumption. For example, when viewing a cooking show on both a tablet and TV, close-ups of food preparation could be displayed on the tablet and the presenter could be displayed on the TV. This enables the user to choose which to focus on while following along. Although some approaches to OBM generation have explored how to generate context based descriptions during capture, they are usually applied in response to pre-defined consumption scenarios such as branching-narrative [91,92]. As such, the type and scope of data required is carefully controlled but lacks the flexibility to produce other types of reconfigurable media at a later date. The task of rebuilding this lost context is thus a more complex and resource intensive process. The existing production pipeline produces a myriad of data/meta-data throughout the workflow, indeed various roles are entirely based around the creation and maintenance of this information. As an example, the *script supervisor*, *continuity editor*, *production assistants* and *camera teams* all use intricate systems of data/meta-data to maintain quality and consistency during a shoot. However this data is: destroyed as part of the post production workflow; and based around tasks rather than context of content of the media.

3.3. Application Scenarios

OBB (Object-based Broadcasting) opens a wealth of opportunities for novel media consumption experiences. To help define the requirements for point-of-capture descriptive data generation, we can think about a number of key production scenarios that can be supported through the production of descriptive and context based data.

3.3.1. New Consumption Scenarios

Cookery is a prime example of a rich consumption scenario with possibilities for OBM production. The kitchen is a constrained scenario that suits sensing in both production and consumption, as such, instrumentation and reconfiguration possibilities are broad and more feasible.

Task Allocation for Multi-User Playback

Playback of a cookery show is typically a linear film with a set of tasks to be completed. In a multi-user playback scenario, each user could be assigned a set of tasks to complete to provide a collaborative cooking effort. To be able to adequately understand where tasks start and finish and which tasks are dependent upon the completion of each other we must have rich data describing each asset. This rich data enables the media to be segmented and assigned to each user, leading to more efficient or playful collaborative playback of the media.

Data required: location of actor, position and usage of ingredients, use of appliances and utensils, what needs to be made before others and blocking points in production (i.e. put into oven).

Adaptive & Customizable Playback Experience

Structured and labelled media affords new opportunities to adapt media to environmental, physical or other constraints, as well as offering the users a means to customize their experience. A cookery show, where precedence and dependencies have been determined allows for a show with multiple components to have sections removed and added dynamically. Simplistic changes such as swapping ingredients and adapting to available equipment in the environment is also possible. The production can adapt to the available screen real-estate, user configuration and other environmental and physical factors, offering a more compelling and appropriate consumption experience for a wider range of devices and environments.

Data required: Segmentation of tasks, description of shot types, identification of ingredients in scene.

3.3.2. New Production Workflows

Semantic labelling of raw media assets at the point of capture affords new opportunities for existing production workflows. I outline a few examples of the application of this descriptive data to the post production workflow.

Sensed Editor

Structure, labelling and organization of media produced from productions is currently limited to those provided by the production team and camera operators. Using data generated during capture I can provide richer semantic labelling on media, facilitating more meaningful queries on corpuses of media. For example, rather than manually searching by thumbnail or preview, a search could be performed using actions or content e.g. “Filleting of the fish.”, “Washing vegetables in sink.”. This offers streamlining of the post-production editing process, enabling quick swapping of shots and the creation of powerful tools for editors to use more natural language querying of media sets.

Data required: segmentation and labelling of individual shots

Quick Editor

The quick editor takes large overwhelming sections of footage from a production and makes simplistic section breaks in order to produce a more approachable and cleaner editing workflow, potentially reducing the ingest time. Initially the system uses luls in captured data to identify the sections where it deems the production was paused, removing breaks in the shoot. Then it can dynamically segment the media from the sensor data based upon continued usage of certain objects labelled sub-sections. This is presented in a chronological order with generated labels describing the utensils, appliances and general area of interest. Editors can preview the generated content from each camera angle and select their preferred shots from each section. Once each desired shot is selected the editor can finalize the edit and the production is cut together.

Data required: media labelled by camera angle, segmentation and content of clip.

3.4. Media of Things

In response to requirements of these example scenarios I developed Media of Things (MoT), a production workflow and supporting infrastructure that is designed to flexibly support the creation of context based data at the point of capture by embedding multiple IoT sensors in a production environment. MoT is designed to capture raw sensor data, in a film shoot context to enable the subsequent creation of OBM as shown in Figure 5.

MoT applies the principles of ambient sensing [90] in which sensors are embedded within the production environment in an unobtrusive manner. This is key to video production where any visual change to the set would not be acceptable. MoT consists of a *'capture'* system for recording raw sensor data from IoT sensors in a production environment, and a *'post-production'* workflow which segments and contextualizes this data alongside any video shot on location.

The nature of OBB being the segmenting and description of objects from raw data streams (such as video, audio and text) means that it is important to capture data in as granular form as possible, so that inferences on the data can be made at a later time as new uses and forms of objects emerge. As such, MoT's primary purpose is to capture time-based raw sensor readings from multiple inputs. Inferences on sensor readings are then made on the data during post production.

MoT is designed to be agnostic to sensor type and volume, and produce segmented data without prior training of specific actions as a starting point for media object creation.

3.5. Capture System

MoT supports many levels of data abstraction and the capture system is designed to reliably record and store native sensor readings from multiple sensors in real-time. This approach to collection of data is in line with other research in data/meta-data capture. Project Orpheus from BBC R&D [110] is a project looking at rich meta-data in radio, in this example the raw input and output from the mixing desk is captured for later processing. This approach of maintaining the original data allows for 'objects' to inferred from the data during post production.

Running hundreds of sensors continuously and reliably within a production scenario (where shooting may not be able to be repeated), with all data correctly time-synced requires a robust system architecture. The system architecture for supporting the continuous streaming of the sensor data was largely informed by current best practices with regards to large IoT sensor deployments as well as similar architecture to the one deployed for Oliver *et al.'s* Ambient Kitchen [72]. A message queue server is at the heart of the system to handle the sheer volume of messages from the sensors. All sensor data is stacked into structured queues to allow listeners to retrieve data for a set of sensors or a specific sensor. The data storage listens on all data sets and stores all data with timestamps for replay. The current system architecture is flexible in that it will accept additional sensors or new sensor types.

A RabbitMQ message queue server running on Ubuntu 16.04 sits at the heart of the Media of Things. A MongoDB instance runs on a separate Ubuntu 16.04 instance with a capture script running to save all data running through the message queue. All sensor nodes were running Universal Windows Apps running on Windows 10 IoT on Raspberry Pi 2.

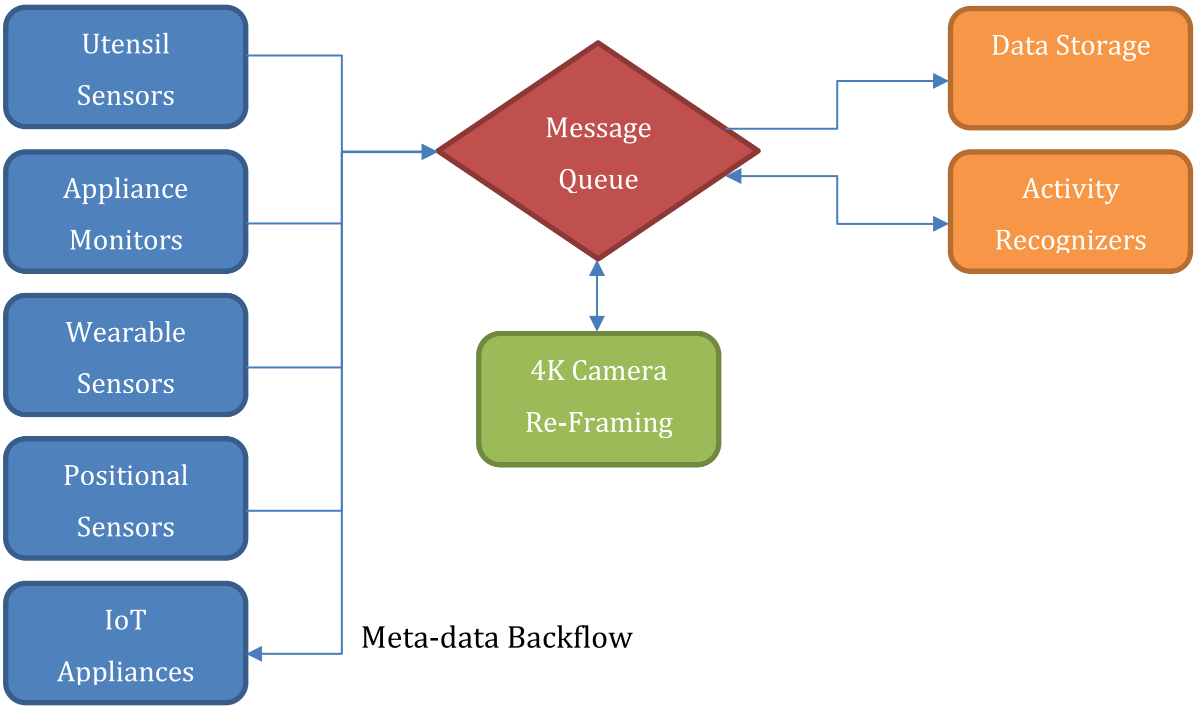


Figure 4: Architecture Diagram

3.6. Inference System

MoT records raw sensor data at a variety of rates from multiple types of sensors. To provide useful inferences for OBB usage, the data is passed through three processing steps.

1. Raw sensor data is converted into consistent discrete time windows and common data format

2. this sensor data is then interpreted into usage information, monitoring when each implement and object is used and when they were last seen.
3. this usage information is mapped against user-defined areas of interest (physical areas of the film-set) generating a probability for each time window of where the focus of the production should be.

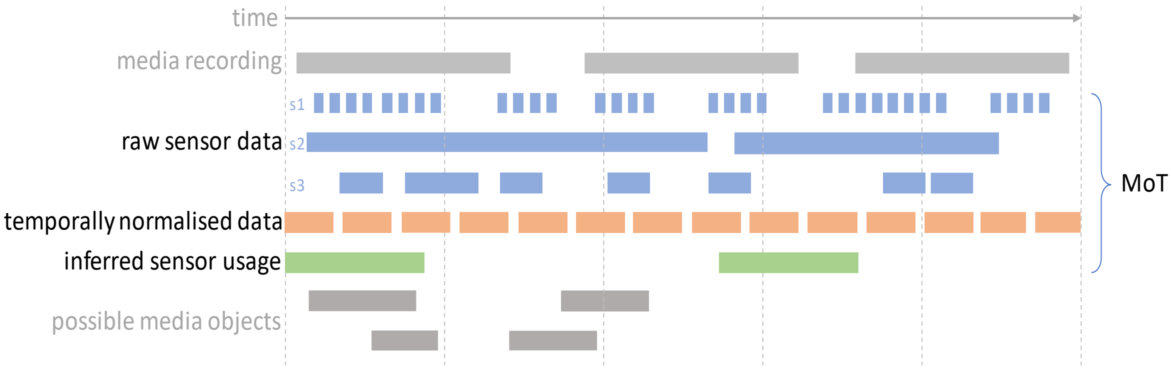


Figure 5. MoT captures and stores time based raw sensor data producing inferences that can be used to create Media Objects

3.6.1. Object Usage Information

Data from sensors are combined into a set of inferences about object usage, including last usage time and duration of current usage. These higher-level inferences provide another level of data from footage, cleaning up noisy raw signal data. This information can be further decomposed into a taxonomy of actions related to each object. As an example, the raw signal from an accelerometer in a knife is constructed into a time-sequenced representation of when the object was last seen, and duration of last/current usage.

3.6.2. Area of Interest Classification

Object usage information from the previous step is combined to produce a probability value for each area of interest in each time window. This value represents the likelihood that this area should become the focus of the production. Degrading significance of pre-configured areas of the production studio are used to maintain the classifications. Sensors have been attributed different significance values, RFID transponders moving or appearing are assigned high significance if they can be located to specific area. Sensors without inherent location (i.e. not RFID) such as utensils are used to maintain the current significance level across the available areas. Appliance usage when located to an area with an RFID transponder is attributed a high significance. This classification could be used to, for example, make decisions on which shot is appropriate in a shot framing system.

I decided to classify 3 main areas of the shoot as a post-production configuration task. This seemed a reasonable compromise in order to improve the quality of the output classifications which would have little implications on, for example, a regular use set.. Starting from a position of knowing nothing about the scene I had a range of devices with different possibilities inferring current areas of live interaction, some devices (e.g. utensil and pots and pans sensors) provide no location information but may indicate continued usage of an area at low latency. Some sensors provide location information but at high latency and sometimes poor accuracy (e.g. RFID and Energy Monitoring).

I will run through an example of the data to explain the function of the area-of-interest classifier. We begin with the classifier being Consider a chef entering our set and preparing some vegetables, they move the chopping board into position on the area to the left of the hob. This is detected by the RFID sensors which is ingested and applied by the area of interest classifier to raise the significance of the left side. This raises the significance above the threshold for a region so is therefore identified as the area of interest. There is a decay in all areas significance over time in order to expire old events and provide more significance to newer data. The chef continues their preparation by using the knife to chop some carrots and coriander. This action is detected by the OpenMovement WAX3 sensors and used to slow the decay of all areas, this was considered so that non-locatable sensors can be assumed to infer that the previously significant areas are still in use. The chef then proceeds to discuss the next step of boiling the carrots in a pot on the hob, this takes roughly 5 minutes with little or no usage of items on the work bench. The current classification of the left area being in use decays over time until it drops below the threshold for significance. The classifier returns to its unknown state. After boiling the carrots the chef then uses a food processor to blend them with coriander. An appliance use is detected with the energy signature monitor and is with an RFID transponder on its base it can be located to a specific area this continued use of a locatable object can be attributed a high significance to a specific area maintaining the significance for that region above the threshold whilst the appliance is in use.

Taking this (rather unprofessional) cookery show I envisaged deploying these classifications to directly drive shot edits. When areas are configured in the view ports of each camera they can be deployed for example in a simplistic way when classification is known frame on area when unknown show a wide shot.

3.7. The Production

As demonstrated by our example scenario, a cookery show is a well-defined scenario in which to test such a production tool. Alongside multiple rich new consumption scenarios, a kitchen is a physically constrained scenario with many opportunities for embedded sensing. In addition, cookery shows maintain a distinct segmentation related to actions on screen, allowing us to test the applicability of our MoT implementation more easily.

Figure 6: 3D Render of kitchen island before construction.

Figure 7: Completed kitchen Island on set.

In partnership with a major broadcaster, we commissioned a cooking programme and hired a team of producers, script writers, lighting, sound and camera crew. The production was designed around Object-based Broadcasting as its intended output and as such the recipe selections were made that would enable reconfiguration with a range of complex technical tasks and utilizing a range of appliances. The production was run with a single chef cooking 4 recipes, chosen as ones with a wide range of technical tasks: requiring chopping, filleting, baking, frying and Bain Marie melting.

3.8. MoT for Cooking

To deploy sensors and IoT devices within the shoot environment which were discrete and largely invisible, we constructed a custom, fully working kitchen island. Although outwardly this looks like a traditional film-set, internally it contains many embedded sensors. To explore which types of sensors would generate adequate information about context, I selected a number of key technologies to offer both redundancy and variety:

3.8.1. Accelerometers

OpenMovement WAX3[28] wireless streaming accelerometers were chosen to capture usage information for utensils, pots and pans. These sensors are small enough to embed without being seen in objects. Accelerometers are a key technology in the fabric of IoT, and provide basic usage information, what utensil or equipment is being moved at any given time. Through a post-production process, this data has the potential to provide rich activity information such as whisking, chopping etc. as demonstrated by Pham and Olivier [76].

In addition, OpenMovement WAX9[29] accelerometer and gyroscopes packages were worn by the chef on each wrist and deployed on the refrigerator and oven doors. In production use, this could be replaced by consumer Smart Watch or Fitness bands. BLE and Zigbee receivers for these sensors were placed inside the kitchen unit and connected via IP to MoT.

3.8.2. BLE Location

Indoor positional information was captured by a LG G Watch[30] placed in the pocket of the chef. This device continuously measured the RSSI of 6 Estimote[31] beacons located strategically around the set.

[28] OpenMovement WAX3, https://github.com/digitalinteraction/openmovement/wiki/WAX3
[29] OpenMovement WAX9, https://github.com/digitalinteraction/openmovement/wiki/WAX9
[30] LG G Watch, http://www.lg.com/uk/smart-watches/lg-W100
[31] Estimote Beacons, https://estimote.com/

Figure 8. MoT Kitchen Sensing: (1) Accelerometers in utensils, pots and pans; 2. RFID positional information on ingredients and cookware; (3) IMU on chef's wrists; (4) Water flow meter on sink tap; (5) Bluetooth beacons for position of chef.

3.8.3. RFID

Benchtop appliances, bowls, plates, and other large items were tagged with RFID labels. Embedded beneath the kitchen worktop were 6 FEIG (0.5m^2) RFID pads[32]. Objects were continuously tracked when on the bench and identifiable from the transponder IDs referenced from a production configuration. RFID transponders are cheap to deploy and come in a range of shapes and sizes to accommodate as many assets as possible. These were placed underneath workbenches, with transponders placed on appliances, cookware and ingredients bowls.

3.8.4. Water Flow Monitor

Water flow meters[33] were placed on the pipework in the sink measuring exact amounts of water dispensed. Since the water system was self-contained (waste and fresh water), energy monitors were also placed on the pump to record usage.

3.8.5. Appliance Usage

Energy signature monitors are widely available on the consumer market. I deployed a Smappee[34] energy signature monitor usage on plug sockets within the kitchen island in turn monitoring appliance usage. Combined with RFID positional information of the appliance, this provides a rich picture of appliance usage.

[32] FEIG UHF Mid Range Reader, http://www.feig.de/en/products/identification/product/id-iscmru102/
[33] Adafruit Liquid Flow Meter, https://www.adafruit.com/product/833
[34] Smappee Smart Energy Montior, https://www.smappee.com/

3.9. Data Player

In order to provide an overview of the dataset and visualise it alongside the collected video I built a data player. This data player pulled together all five camera feeds and each of the raw datasets time synced and visualised on graphs and other appropriate data representations. Each item is labelled according to its production configuration providing more useful names like “serving spoon” and “frying pan”.

3.9.1. Implementation

BBC Primer (used to record the production) is an IP and web based media capture suite. The Media of Things data capture suite was also an IP based sensing system, as such I decided to build the data player in a web browser. Web technologies are easily deployable and platform agnostic (largely), also BBC Primer records its content in a “dashed” format, in this case MPEG-DASH[35]. MPEG-DASH has a much lower overhead for webservers than traditional server block-based data retrieval, content is segmented into separate files and retrieved based upon a calculated file number using the current time (live) or historic time (on-demand).

In order to sync the film from each camera and the data I used TimingObject[36], this timing object provides a single truth of timing, when used in conjunction with video elements in a web page it drives elements forwards and backwards by adjusting the playback speed. This ensures that all films remain in sync, and timing updates are provided to data graphs to ensure sync too.

The viewer and all static data are served from ASP.NET Core MVC[37], all client-side scripts are written in TypeScript[38] and transpiled to JavaScript for running in the browser.

---

[35] MPEG-DASH, https://en.wikipedia.org/wiki/Dynamic_Adaptive_Streaming_over_HTTP
[36] Timing Object, http://webtiming.github.io/timingobject/
[37] ASP.NET Core MVC, https://docs.microsoft.com/en-us/aspnet/core/
[38] TypeScript, https://www.typescriptlang.org

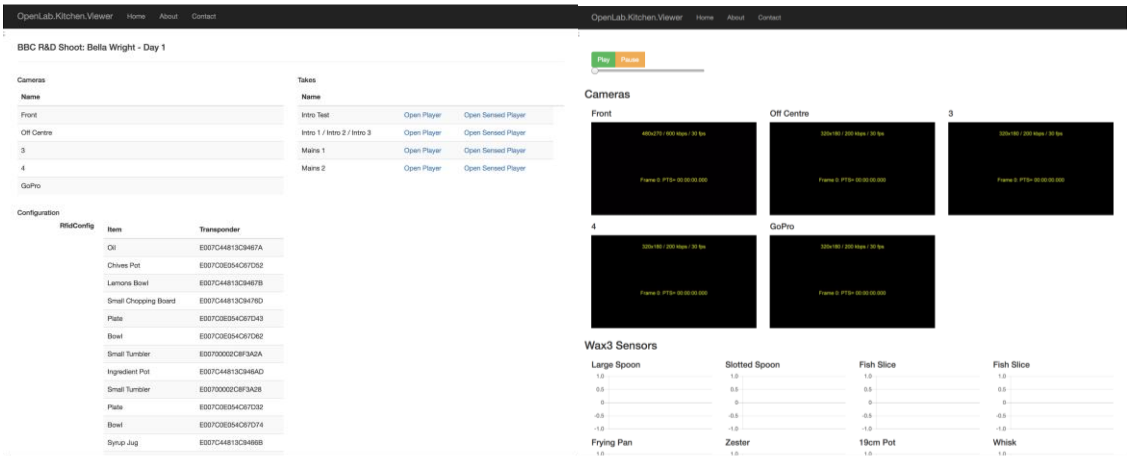


Figure 9: Production configuration screen (left), Data player with 5 camera feeds (right).

Data accompanying the media is served from a ASP.NET Core Web API instance and is retrieved in time ranges from the server. All requests are proxied through a simple client proxy which caches around the requested timeframes and caches all requests to reduce the number of requests made to the server and present the data faster. WAX3 Sensor data is presented in graphs as shown in Figure 9. RFID transponders are shown as red/green squares indicating detected presence and labelled with which pad they are on. Appliance data is similarly represented with red/green squares depending upon whether they are detected off or on respectively.

3.10. The Film Shoot

The production took place over three days in a black box studio location in 2016. The crew consisted of ~20 professionals including lighting, set, electrician, camera crew, director and producer and talent who were hired for the production. A TV chef was employed to perform the preparation of 6 meals on set, which were captured by both static and manned cameras over the production. In total, we captured ~12 hours of sensor data using MoT and continuous video footage from four Panasonic DMC-GH4[39] using BBC Primer [11] for time synchronization and capture from the production over two days of shooting. To aid with data validation and to enable development of future tools, we deployed an un-manned 4K static camera system consisting of four wide-shot cameras around the set. These were capturing continuously during the shoot. MoT was deployed to capture raw sensor data during this same period. Around 72,000 sensor

[39] Panasonic DMC-GH4 4K Camera, https://www.panasonic.com/uk/consumer/cameras-camcorders/lumix-g-compact-system-cameras/dmc-gh4rh.html

data points were captured alongside four camera angles. Three members of our research team joined the production team to manage the equipment, data capture and physical components of the augmented set.

3.11. Production Report

The production was an overall success in terms of data and video captured, however a number of incidents and specific failures during the shoot offer us a chance to reflect on how MoT performed within this context.

Whilst the production was planned with the whole production team being aware of the focus on OBB, the production workflow failed to adequately adapt. The need for sensors within the set formed a limited part of the production planning resulting in utensils and cooking equipment being hard to augment with sensors once on location. RFID transponders provided reliable object location information throughout the production. However due to the equipment being washed between shooting sequences RFID tags were sometimes accidentally removed from devices, requiring the researchers to re-apply and record tags. Additionally, the material of some objects (such as glass bowls) prevented tags being attached at all.

During the production, the WAX9 sensors (to be used for gestural detection for the chef) unfortunately failed to function during the production due to a data transfer issue, resulting in malformed data. Additionally, a laptop being used as a data capture relay overheated under the studio lights resulting in the data being lost for a take early in the shoot and before the start of the production the water flow meter failed. However, the energy signature monitor offered us redundancy for tap usage data.

Despite failures, the built-in redundancy and reliability of other sensors led to a good coverage of actions being performed within the production. The success of the data collection largely vindicated our choice of sensors for this type of media production. In particular, the WAX3 accelerometer data produced a consistent and useable dataset.

Very quickly, whilst I was operating the MoT system, I became integrated into the production team's workflow, and roll-call. Temporary sensor failures and resets became analogous to lighting adjustments and camera focus issues to the production team, and I was considered an equal member of the production team who was able to call 'halt' when required for adjustments. This enabled many temporary issues to be addressed immediately resulting in a more consistent dataset. However, understandably we were only able to use this relationship in a limited fashion, as maintaining the production schedule was of primary importance.

In total, data from 25 kitchen implements (i.e. utensils, pots and pans) was recorded, alongside RFID position information for 28 objects (i.e. plates, bowls and ingredients), and power usage data from three appliances. After performing MoT's post production inference process, the data consisted of ~450 discrete movement actions, and ~200 changes in positional information for objects on the kitchen unit. The output of which was time-based data related to individual tagged items on set. These were to be mapped against the production notes made by the team to determine which utensil or item was used.

3.12. Design Recommendations

The deployment of MoT integrated within an existing production workflow provides us with invaluable insights into the suitability of existing production. Through self-reflection by the our team and observational notes of the shoot on our deployment and the challenges and successes of integrating MoT within a film production workflow, I can draw out three key design recommendations for OBB context capture:

3.12.1. Introduction of the 'Sensor Operator'

The introduction of new technologies and workflows into film is a slow and difficult process, due in part to the inherent link between the experience of film crews and the technology they use. The industry's response to the introduction of new workflows has historically been to apply a new team role, as seen with the introduction of digital film media (the 'DT' role), and computer graphics (the 'CG' role) which fits within the well understood practices of the team. It is this standard practice that enabled me to build legitimacy within the film crew as a member of the production team, thus providing the ability to request pauses of the shoot (through negative confirmation when "ready sensors" was shouted) to reconfigure or fix emergent problems. Being physically located on set also increased recognition of the process to the rest of the crew, which led to easier integration of production tasks such as resetting objects on set for continuity. Given that MoT can provide live feedback of sensor inferences and instant recall of time-based data, I envision this tool could be used by the production team for more efficient continuity adjustments on set (e.g. we could detect an item is missing from the scene).

3.12.2. Embedded Sensors in Production Environments

Our deployment clearly highlights the importance of redundancy in recording sensor data in production scenarios. However, the selection of sensors used to capture is more nuanced and there are multiple factors which contribute to this selection: visibility on camera; ability to embed in the environment; amount of pre-configuration required; and access for adjustment and the quality of data captured.

In the context of a kitchen, where the set is physically constrained and where a limited known set of manipulability objects exists, accelerometers proved to be a good balance between visibility and rich data, and provided useful information. I felt however that the use of BLE indoor localization technology did not justify the potential results. To obtain sufficient granularity of position, the sensor technician had to perform a lengthy (~2 hour) calibration of the sensor before it could be used, and the resulting data was only sufficient to identify which end of the set the actor was located.

In many cases (such as with raw food), objects are unable to be directly tagged, so proxy tags are used (such as the packet the food came from), but this has limitations, particularly when dealing with fresh food. Utilizing customized embedded vision approaches such as FoodBoard [75] or IRIS [64] a prototype imaging surface I work on with my colleagues would be one way of obtaining this data. These systems provide possibilities to image items on the surface and apply rudimentary object recognition from shape and/or colour to identify items. Techniques such as this are also a possibility to reduce instrumentation of individual objects instead relying on vision based recognition of objects.

Given the expensive and temporally compressed nature of film production however it is important to be resilient to sensor failure, as there may not be a second chance for capture. In our scenario for example, RFID pad coverage overlapped, and water flow information was available through pump usage on the energy signature monitor. Therefore, MoT could not only cross-validate sensor data but support multiple points of failure.

Key in supporting this is the Sensor Operator's ability to watch a live feed of all sensors' data being recorded by MoT allowing them to respond appropriately when issues occur, just in the same way that the sound operator monitors his recording. Such an interface, if present in our production, could have enabled us to detect errors and fix them before the cameras rolled.

3.12.3. Configuration as a Key Pre-Production Task

Although MoT is both sensor agnostic and context agnostic and thus does not require pre-training or ground truth data in order to infer data, each type of sensor may require configuration on location. In particular, sensors need to be recorded against a production reference that will make sense in post-production (e.g. 'the spoon').

Neglecting to sufficiently plan the film shoot with these tasks and constraints in mind led to confusion during setup of the shoot, as set design, production managers, and sensor operators struggled to find a solution for problems that should have been identified earlier.

Each sensor required careful labelling, and each transponder needed to be mapped to an object. The process of this configuration was very time consuming as a record must be kept of which sensor is assigned to which object, while this can be recovered manually post-production from the footage. This experience demonstrates that sensing needs to be considered at all points of the pre-production pipeline. The type, location and variety of sensors should be considered alongside the design of the set, and production timings should be adjusted to include sensor configuration and setup. In hindsight, a shared codebook of set objects should have been shared amongst the production team, enabling a shared reference for labelling objects during production changes. Future productions should also consider the importance of labelling individual objects. Each new object added to the production as a labelled artefact or member increases the complexity of the configuration process.

3.13. Making use of MoT Inferences

MoT is designed to facilitate the production of sensor agnostic descriptive data that can be used to infer specific context throughout the production and consumption pipeline. The raw data and basic inferences about object usage within the shooting environment can potentially be used to drive new workflows, as envisioned earlier in this thesis. This generic and fundamentally reusable approach is a shift from existing work in OBM production which requires the output scenario to be pre-defined. I believe this to be an important step towards enabling a sustainable workflow for data generation which can be applied without re-configuring well-defined professional workflows.

Although useful to broadcasters and top-down media production, I propose that MoT opens up important opportunities outside of professional production. Much of current IoT development is consumer product driven, thus by approaching descriptive data generation with IoT as a possible solution I have demonstrated that products already within the consumer environment (e.g. RFID tags present in many home goods) could form a key part of the media consumption and production experience. Indeed, one could imagine how IoT appliances could be used to record rich descriptive data whilst recording home YouTube cooking videos, or to enrich the viewing of immersive content by adjusting ambient lighting or audio devices.

The workflows and processes I have developed around MoT go towards generating raw and low-level inferred data. However, I acknowledge that this data, without context applied by the production team during post-production, is not sufficient to define fully formed context-based objects that could be used directly by a consumption scenario. There is a significant creative process that directors, producers and designers must apply during the rest of the media production pipeline – how a playback experience utilises this data to create an immersive experience.

3.14. Summary

In this chapter, I present Media of Things, our implementation of a point-of-capture data capture system which integrates into the existing production workflow. By supporting the capture of raw, time-based sensor data from multiple sources in real-time, MoT provides a solid base for creating rich Object-based Media experiences in the future. I highlight the symmetry in using IoT based technologies with 'invisible' sensors which can record actions and object use on a film set, responding to the need to capture such contextual and descriptive information from broadcasts in nearly-live scenarios and post production. Through the deployment of Media of Things in a real production scenario, I explore how -data capture can become part of the film production workflow.

The primary contribution is the validation of sensor-based meta-data capture as a sustainable and flexible process that can be included reasonably within the existing constraints of the media production pipeline. Specifically, I recommend that the 'Sensor Role' should be recognized as a distinct and valued member of the production team, in line with the roles allocated to 'Sound' and 'Lighting' professionals, facilitating a smooth transition into the professional workflow. When planning for MoT style productions however I caution that the selection of sensors is a nuanced and often a difficult trade-off between granularity of data and operational requirements for configuration. This is an area that I acknowledge would benefit from more work, trialling different sensors for the collection of data.

Although our deployment was in the constrained environment of a cookery show, the structured and standardized nature of film production means that MoT has been used against the inherent situatinal factors which define such workflows. Combined with the flexibility offered by our sensor agnostic approach, I envisage that MoT could be used as part of the normal production workflow for any production scenario, when appropriate sensing is applied. In summary, I encourage the media research community to consider how production tools such as MoT, leveraging developments in IoT, can be integrated into existing production pipelines now, to more rapidly enable the rich, immersive and indeed exciting experiences that Object-based Broadcasting can deliver.

In elucidating the motivation behind this research, I discuss the possibility of developing greater detail in labelling from media productions. The focus of this thesis is to enable content makers to design for the unknown and to demonstrate mechanisms through which this can be achieved. Conceptually, the Internet of Things shares similar goals in terms of providing new sources of information in order to construct a more detailed picture of a situation, or greater control over an environment. Deploying such technologies within a production proved fruitful. Much of the related expense was due to upfront costs and the developed system and output is able to support a range of consumption experiences, with adequate continued development. In labelling the media assets at such a granular level I have greatly developed the reapplicability of the media, labelling which actions are being performed and, in some cases, their rough region within the camera viewport. Reconfiguration of the media can now be performed based on a wider set of parameters and inputs. Crucially this was achieved with little impact on the production workflow, indeed with greater planning I believe the sensing impact could be reduced even further. Much of the time and expense were upfront build costs, mitigated by Open Sourcing our works for future research/productions. Sensing in production environments has proved successful in the aim of providing more detailed labelling of media assets.

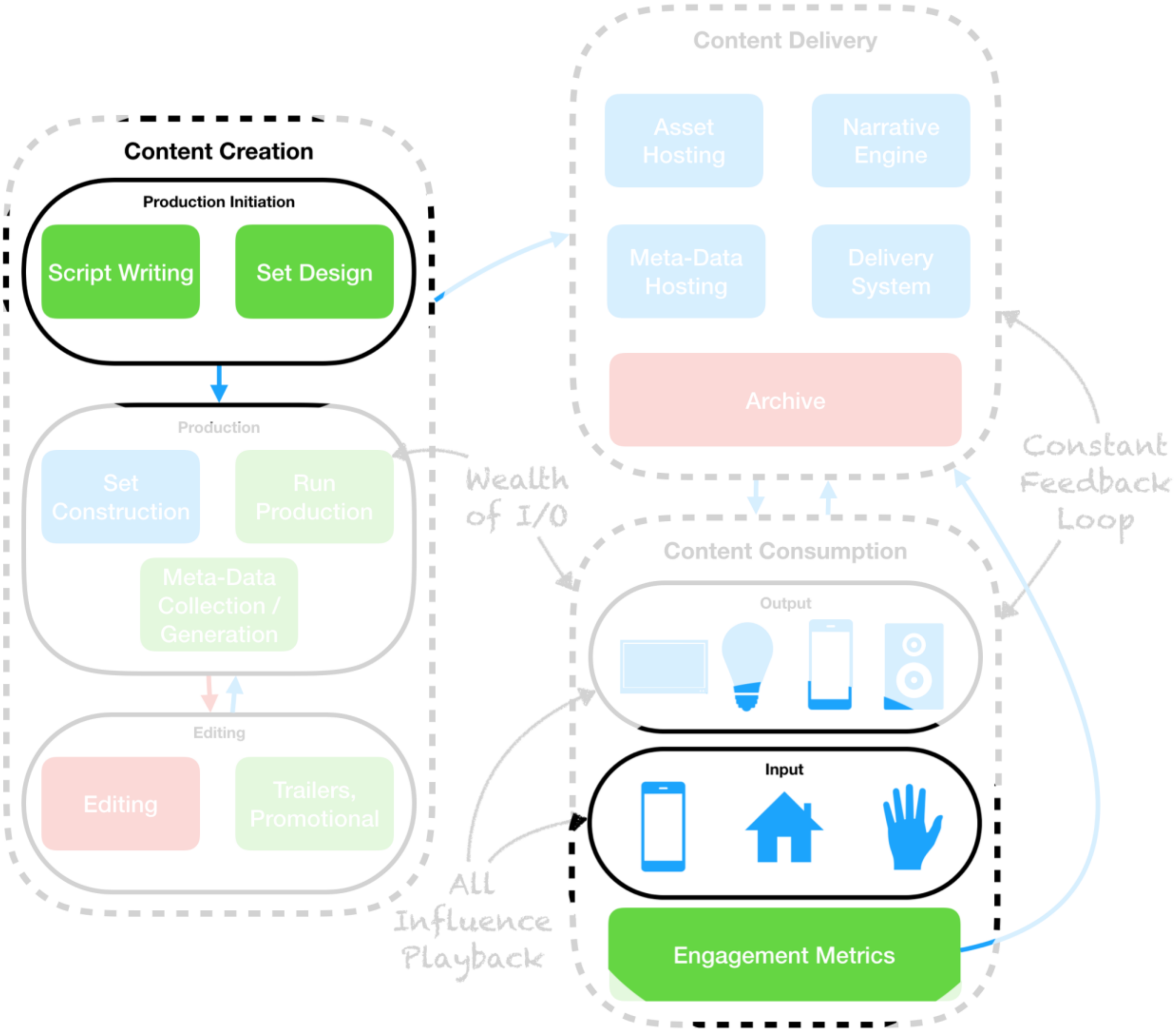
The research presented in this chapter resulted in a sensor-instrumented portable kitchen, and a large dataset of sensor data and film (Table 3 surmises the dataset captured below). I refer to this Media of Things dataset throughout my thesis, I used it to focus my research. In doing so, I wished to further validate sensors in productions, and produce working prototypes where appropriate and necessary. Moving forward into Chapter four, Integration, I explore the opportunities for greater understanding from existing datasets – personally captured, infrastructural or otherwise. In doing so, I look to develop technical solutions which could facilitate the generation of OBM for consumers and create even richer datasets from professional productions. I also investigate how existing descriptive data generation tools have failed to realise Object-based Media as a pervasive consumption technology.

Table 3: Summary of Media of Things dataset.

Sensor	**Data Type**	**Resultant Data**
WAX3 in Utensils	3-axis accelerometer readings (50Hz) – time-stamped	Movement and action data (with taxonomy) of each utensil
WAX3 in Pots & Pans	3-axis accelerometer readings (50Hz) – time-stamped	Movement and action data (with taxonomy) of each pot/pan
RFID on Appliances	Transponder ID on Pad ID – time-stamped	Location of each appliance on work bench, accurate to defined regions
RFID on Cookware	Transponder ID on Pad ID – time-stamped	Location of each item of cookware, ingredients in use
WAX9 on Chef's Wrists	3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer (50Hz) – time-stamped	Conducted gestures, subtract from other sensor data to reduce dataset noise
Energy Usage of Appliances	Watt-change, appliance ID – time-stamped	Appliance usage
Pre-defined Camera Shots	x,y coordinates of 3 pre-configured areas	Potential re-framing options
Location of Chef on set	Bluetooth RSSI data – time-stamped	Fingerprint location information of Chef along workbench

Integration

Chapter Four



Open Lab at Newcastle University

BBC

Chapter 4. Integration

Data/meta-data has always formed a major component of media productions, scripts, set designs, running orders and metrics. These are vital components of the exercise of media production, however these assets are often kept in siloed systems or in physical form, limiting their possibilities for integration into other parts of the production workflow. Similarly, the wealth of personal data and environmental data we collect is typically viewed without adequate contextualisation, for example causes of: increases in heart rate; disturbed sleep; and spikes in air pollution in the area.

In order to realise the full potential of Object-based media I will investigate how these components can be re-integrated into the production workflow, providing the necessary contextual information to enable OBB. Moving beyond productions, I also look at the opportunity for greater inferences from data we already gather about ourselves. In doing so, I aim to provide compelling use cases beyond typical use cases with deeper integration of services, maintaining raw data feeds to the user and propose more open approaches to data sharing.

In this chapter, I look to explore deeper integration of datasets. Hawk-Eye EDGE-SENSE was deployed as an example of sensing in sports with an open dataset captured and used for officiating. Investigating the possibilities for the quantified-self movement [79] data to be used as an input for officiating and professional broadcast services. CryptoCam looked at a means for passive collection of video footage, much like passive data collection from a fitness tracker or sensor. The subsequent debate and discussion provoked by this technology informs the future development of greater integration of content and services.

4.1. Production Application of Descriptive Data

As outlined in the chapter Related Work, data/meta-data is already widely deployed in productions for visualisations, take labelling on cameras, script markers and audio descriptions. If we take a typical editorial process for a professional TV show: during the production there are a wealth of processes and undertakings to assist editors, from more fundamental narrative roles like continuity editors, script and shot supervisors; to more subtle assistive processes like camera/sound operators labelling media assets with scene markers, re-take markers and more. Once the cameras are switched off this information is exploited to expedite the editing process, however the process is still very much manual.

Advocates for a more integrated approach to content production have long been calling for changes to the production process, to provide a more integrated editing process. Indeed, Davis *et al.* [23] back in 2003 was arguing that one of the most significant developments to assist in democratisation of media creation was in redesigning the production workflow to be less wasteful. Integration of these existing datasets, produced in professional productions but destroyed, or not maintained beyond the end of the production or the editing process was key. The paper discusses the movement from physical film and while much of the content is dated, the sentiment is generally content productions should be a more integrated process. This would, I believe, enable a greater range of assistive tools for editors, both in terms of better indexing for the wealth of available media – but also for moving beyond this into the realms of assisted editors. Suggesting substitute clips based upon detected similarities in descriptive data as well as framing.

4.2. An Editor's Paradise

In the chapter Media of Things, I focussed upon the opportunities for sensing in relation to content generation with Object-based experiences. However, there are opportunities for improving the editing experience too. Editorially, Object-based Broadcasting presents a whole host of potential headaches. Introducing reconfigurability and responsiveness to the output inherently increases the underlying complexity of the editorial process. New variables have been introduced, much of them potentially inflicting fundamental changes upon the media presentation and narrative. This leaves editors with a largely unbounded set of possibilities when it comes to curating an experience for users. Not even the set of experiences is defined, when it comes to a true Object-based Broadcast. Tools must be developed to assist in this process, developing the complex logic models backing these experiences.

However, on a more superficial level this new depth in data labelling provides a potential boon to develop novel assistive/automated editorial tools. Consider for example, a typical edit to remove a shot where a chef in a cookery show has made mistake in preparing some vegetables. Finding a replacement shot would typically involve manually finding the replacement footage, potentially sifting through different asset libraries or different days of shooting to find the right content. Rich data labelling could however be used to assist in edits like this. Matching similar clips with similar data/meta-data we can suggest replacement clips using this data as a means to index media in an incredibly nuanced and detailed manner. In the example above, the knife movement could be tracked along with the ingredients and chopping board on the bench. Matching clips with the same tracked features would then be a trivial task for an assisted editor. Moving beyond this research into advances in Artificial Intelligence and increased levels of descriptive data could provide even more in-depth assistive editing tools. BBC R&D have been working, for some time, on an Artificial Intelligence Producer [111], investigating applying Computer Vision techniques to segue between different camera angles and reframing shots. At a higher level, this could be an assistive editor rather than explicitly automatic, suggested shots using vision – backed up by descriptive data – could prove a compelling editorial experience.

Through deeper integration of existing media production datasets, I believe we can create a deep creative environment for content editors when producing existing linear content and new Object-based experiences. Works like Storyarc [112] are providing data models and tools to develop narrative engines and logic models backing these complex Object-based experiences, but on a more superficial level increased descriptive data provides a rich and fertile environment for editors and 'smart' editorial tools.

4.3. Information Siloes

Object-based Broadcasting as a concept inherently relies on a wealth of contextual information on relevant media assets. In these experiences, logic models can dictate the playback of the media, deciding which assets are played, in what order and through which medium. Crucial to these logic models is the descriptive data fed into them, enabling them to make logical decisions to output media.

In a typical media production there are a number of systems involved pre-, during and post-production:

- Scripts are developed for each part each labelled with scene information and contextual hints relating to composition of the talent.
- Set designs are developed, perhaps with 3D models of parts of the set, carefully positioned camera information and relevant entry and exit points.
- Camera feeds are captured into a media capture suite, meta-data describing the camera the feed came from, perhaps a shot description, time stamping and marks to denote current scene and reshoots.
- Editors and Producers may generate labels from mixing desks as a production unfolds, marking areas which are of note and making initial editorial decisions.
- In live productions, the very sequencing of the shots is performed from the mixing desks.
- Editorial processes label and remove segments of the raw media assets which are surplus to requirements, some of these processes look to directly re-associate certain assets with parts of the script, reconstructing the narrative.

In each example above, traditional media productions store and process these datasets in siloed systems. Meaning that cross referencing and longitudinal exploitation of this data is limited. This is an important consideration moving forwards, how can we ensure that we can make the best use of pre-existing recorded contextual information before we should even consider generation of new descriptive data. Standardisation of supplementary data storage with systems like BBC's IP Studio [113] demonstrate an industry interest towards bringing this information together. Encapsulating this information in a more open and organised manner, enabling linking of assets together, we can make best use of this data enabling OBB logic models to make more complex playback decisions.

An obvious limitation in any such system integration is the inherent destructiveness of some parts of the production process. Often this is a necessary step, removing reshoots and segments which are cut from the final output. However, some editorial decisions, framing of the shot or cutting to a different camera ought to be reversible. Systems such as BBC Primer [11] perform shot reframing virtually, piping out the reframed shot for broadcast but maintaining the original film. These approaches are inherently conservative in nature, ensuring that the destructiveness associated with film making is limited while also limiting impact on production workflows.

4.4. Data in Sports

Measured excellence and marginal gains have always been at the forefront of modern professional sports. Team training camps in Cycling, Rugby and Football track information about their athletes down to the minutia of detail. This information is fed back into performance monitoring and improvement calculations. Often posited as the reason for teams success in the past [114], the contribution of these stats to coaches and overall performance is undeniable. Broadcasters have also sought to provide more compelling experiences to their audiences through collecting and presenting detailed match/event statistics tracked often in real-time. Much like many of Object-based Broadcasting examples above many of these metrics are manually generated, particularly in football (Opta[40]). The BBC have experimented with graphical elements of a show being rendered independently of the footage, allowing elements to be resized, moved or not shown [54]. While this information is captured for broadcast it is rarely presented alongside the broadcast media in a machine-readable format. As such its post-production application is limited. In the EDGE-SENSE project, I provided raw sensor data to broadcasters alongside the umpire/detected decision systems.

4.4.1. Hawk-Eye EDGE-SENSE

In professional and non-professional cricket the subject of being caught out for an edge is a contentious subject. As with Football goal-line technology we look to technical solutions to eliminate this kind of ‘cheating’ from our sports. In each case, companies such as Hawk-Eye look to track aspects of the sport for later review. This meta-data alongside the film is then deployed to provide evidence of mistakes by officials and to generally encourage better behaviour from competitors. Snick detection in cricket has been approached through a number of mechanisms though mainly with thermal cameras (Hot Spot[41]) and high-fidelity microphones (Hawk-Eye ULTRAEDGE[42]). While these systems provide reliable evidence to officials of the professional game, they lack practicality and convenience to be more widely adopted at club level cricket for example, not to mention the hefty price tags associated with most systems.

---

[40] Opta Sports, https://www.optasports.com
[41] BBG Sports – Hot Spot, http://bbgsports.com/
[42] Hawk-Eye – DRS, https://hawkeyeinnovations.com/products/ball-tracking/cricket-decision-review-system

As part of my research into data/meta-data in existing media productions, and how it is exploited, I worked on an exciting collaboration between Open Lab and Hawk-Eye Innovations[43]. I produced an iPad application that would, when connected to an OpenMovement WAX9 IMU[44] on a cricket bat, allow umpires to see the faintest of 'edges' from the ball when caught out.

Implementation

The application ran on an iPad situated near the umpire pointing down the wicket. The device would record in its highest frame-rate (in order to provide adequate slow-mo footage) on a 30 second loop. The sensor (WAX9 OpenMovement IMU[44]) was located on the back of the cricket bat, connected over Bluetooth to the iPad/iPhone streaming data back with roughly 100ms delay. This data was also captured and stored in a 30 second buffer for later review. Upon a contentious decision the umpire could press "Review" on the screen which would begin the review process. In Figure 10 you can see the review screen with a small snick detected. The 3 axes of the sensor are shown on screen to the umpire along with the video replay. The white line down the middle denotes the video sensor data sync point. The umpire could scrub left to right across the screen to move the video forwards and backwards, and play the video in slow motion. After review, the footage can be marked as IN/OUT with buttons at the top. The video files and sensor data are marked based upon the decision and stored on the device. At the point of review the video files and sensor data are sent to Hawk-Eye's Smart system for broadcasters to use in a professional sporting event production.

[43] Hawk-Eye Innovations, https://hawkeyeinnovations.com
[44] OpenMovement WAX9, https://github.com/digitalinteraction/openmovement/wiki/WAX9

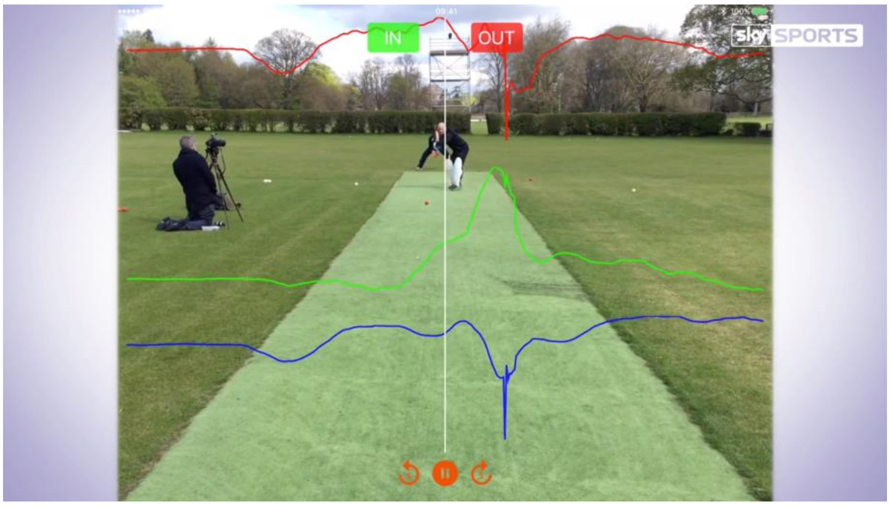


Figure 10: Hawk-Eye EDGE-SENSE application from Sky Sports.

This technology has application in both club and professional levels of cricket and was showcased in The Times [44] and Sky Sports News [85]. Crucially, I maintained raw sensor data throughout the review process, the signal data and footage are sent to broadcast systems for rendering or perhaps broadcasting raw to end users. Such an approach enables the greatest range of re-applicability and deeper integration with other content and services.

4.4.2. Integrating Sports Data

There is a wealth of opportunities for greater integration of these types of datasets. Typically, broadcasters track and present an enormous amount of data to the consumer in graphical overlays. However, this data is often not maintained or broadcast in a machine readable or raw format. Indeed works from the BBC with Forecaster [54] experimented with fragmenting assets of the broadcast media in order to allow for reconfiguration at the client side however I argue this needs to be at a lower level. In production editing media assets filtering based upon statistics such as: goal scored; track position in Formula 1; or heart rate of riders in the Tour de France, could provide a more efficient editorial experience. Similarly, in consumption environments media could be dynamically reconfigured, exploiting the same datasets to create highlights to suit the user's duration requirements, or create ambient effects with lighting in the home based upon track position in Formula 1.

4.5. AI Producer

Artificial Intelligence in film is a steadily growing industry and subject of research. Films have depicted robots taking over the world for years, often with nefarious motives and typically disastrous effects. The media industry are looking at applications for Artificial Intelligence in all aspects of the production workflow from copyright infringement [89] to script writers [118].

Moreover, there has been a wealth of work in automating aspects of the production and shot selection itself as outlined in my Related Work. In the Media of Things chapter, I detailed one of the meta-data outputs as an Area of Interest. It classified which part of the film was the most salient at any one point in time. Currently, the logic model backing this classification is highly specific to the sensing and production context. In the case of the cookery production I defined areas of interest based on accepted cookery show shot practices (face-up, detail and presentation shots). As such, I targeted sensors to detect these shots: RFID pads on the front portion of the bench; and sensors on the chef's wrists. Using AI, even in its most basic of senses, with simple logic models governing when such areas are the most salient, I could construct an AI Producer to guide the cameras in a similar manner. Applying principles of cinematography to these areas of footage I could create an auto-director that provides a satisfactory coverage of a constrained subject such as cookery show. Crucially, I believe explicit sensing provides the required contextual information to enable AI Producers to function adequately, like those proposed by Kaiser in Virtual-Director [50].

4.6. Consumer Generated Data

In this modern age of measured excellence, individuals and organisations collect a wealth of information, from health metrics to consumption statistics – technology companies are motioning to exploit this information in new and innovative ways. HealthKit[45] on Apple iOS looks to coordinate the secure storage and sharing of such information. We view this information often in a largely siloed context, how much exercise have I done today? What was my max heart rate? These insights, while useful and informative, provide a limited overview of how a user's day occurred.

[45] Apple HomeKit, https://apple.com/ios/homekit

Thus far in this thesis, I have focussed on professional collection and application of data/meta-data, however this wealth of data generated by users could be exploited to create Object-based Media. Media production is constantly evolving too, independent production companies and individuals are seeing a huge surge in audience and funding. YouTube largely drives this medium with funding through projects like YouTube RED[46] and advertisement driven services. When considering Object-based Media, we need to ensure that this type of content generation is also available to consumers/independent studios. As such, I looked to explore a means to introduce rich descriptive data into consumer production workflows.

4.6.1. Indexing Film with Data

Constructing films from existing media based on rich data could prove a compelling experience for users. Not only can we exploit the wealth of existing data collected but at the same time reduce potential costs both in terms of time and money, related to the generation of Object-based Media. In researching this area, I discovered an enjoyment for the concept of a camera as a sensor. Cameras, after all, are one of the most powerful sensing devices ever produced, however they are difficult to adequately maintain privacy within most contexts. As such, their application in the smart home and workplace is restricted to security and protection. I sought to develop a technical solution to this perception problem with cameras. I constructed CryptoCam, a privacy conscious camera with secure sharing of footage to subjects nearby. CryptoCam is a conceptual camera network with a more open approach to media sharing. In the paper [99] we explore the opportunities for sharing access to footage from CCTV cameras to potential subjects nearby.

4.7. CryptoCam

Video cameras are versatile tools, used for a plethora of purposes, from capturing one's day to day life (lifelogging), as an assistive technology [40,58], a means for sousveillance [57], as a tool for documenting human rights abuses [31], as a scientific tool [105], as a form of sensing (e.g. Kinect[47]), for recording current affairs (for journalism), or as a tool for enhancing security in public spaces. Yet, many existing surveillance systems are

[46] YouTube RED, https://youtube.com/red
[47] Microsoft Kinect, https://developer.microsoft.com/en-us/windows/kinect

closed in nature: control often resides with the organization which has recorded the footage, or to put it another way, with the watcher rather than the watched. For the ordinary citizen, the process for obtaining footage can be challenging. One would have to notice the camera (many are obscured), determine who owns it and in turn, make a formal request through legalistic processes. The result might be that someone is provided with the footage several months later on a DVD (in a format and timing that might well be of little use).

The closed nature of surveillance cameras, along with other privacy challenges, for example that video footage can sometimes contain (directly or indirectly) sensitive personal data, and the desire for individual privacy more generally, leads to concern over their configuration. There is a pressing need for cameras, but also a similar need to configure these systems in a way that does not overly intrude into the rights of those being recorded. Researchers have adopted a variety of approaches aimed at achieving this, including careful positioning of cameras to ensure that their field of view is only focused on the target of interest [87]. However, discoverability is an under explored area for CCTV, in terms of presence, purpose and configuration.

CryptoCam looks to enable new possibilities for configuration by deploying industry standard encryption technologies to encrypt footage at point of capture, with keys to the footage distributed to relevant parties. The emphasis of CryptoCam is in redressing the asymmetry of control and access over footage from cameras. Through moving the locus of control from operators to potential subjects, either in its entirety with a complete local encryption solution, with keys only held by subjects, or a more balanced solution, with master keys held by the camera operator. A client application on subjects' mobile phone records keys from cameras nearby using Bluetooth, placement and other configuration information can be shared at this time. The process is entirely anonymous with no identification of the subjects needed or warranted. This particular configuration has several advantages: in many settings (within the EU), it would fall within the domestic use exemption of the General Data Protection Regulation (GDPR) it is likely to enhance user trust (especially in the context of negative perceptions of cameras). I hope its use will be accepted in a broader range of scenarios, while also enhancing consent and accessibility in relation to the footage.

4.7.1. Technical Implementation

CryptoCam is a prototype Open Circuit Television [99] system, using state of the art encryption technologies and Bluetooth to secure footage and share access tokens with subjects. Combining existing technologies to produce a simple yet powerful system to realize the possibilities for Open Circuit Television.

CryptoCam is a camera that combines encryption and Bluetooth to create a secure and anonymous footage sharing system. Video footage is taken from the camera and encrypted, the key used for encryption is randomly selected at the beginning of the recording interval. This key is made available to users nearby over Bluetooth (Figure 12). Software clients installed on users' devices listen for encryption keys broadcast from cameras nearby, storing these keys for possible later use. The flow of data is described in Figure 11. Cameras also describe a file access protocol; these files can be later retrieved in an encrypted form from the specified location. If required, the encryption keys collected locally are then used to decrypt the encrypted video footage once retrieved. Crucially, all decryption occurs on the device, the key never leaves the device so the user is not identified to the server and the encryption key stays private to the camera and nearby devices. At the end of a recording interval the encrypted recording is uploaded to the specified file store, erased locally, and a new encrypted recording is started using another, randomly selected, key.

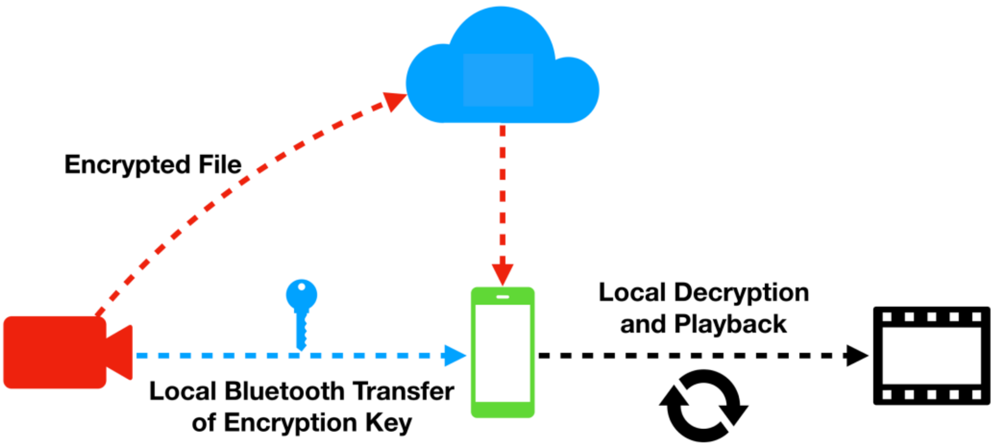


Figure 11: Footage is recorded, encrypted and uploaded to a cloud storage provider. Encryption keys are distributed to phone-based listening clients nearby. Phone clients later retrieve encrypted footage from the cloud and decrypt the contents using the key previously provided locally, then the footage can be played on the device.

Camera

The prototype implementation uses a Raspberry Pi Zero W[48] with a Pi Camera v2.1 module. This hardware provides an inexpensive, small and low power package which is highly configurable. The CryptoCam software runs on Node.js[49] and an industry-standard encryption stack (OpenSSL[50]) is used to handle encryption of footage and generation of keys. The camera is configured to continuously record with a pre-configured segment interval. A random 256-bit key is generated and broadcast to any devices listening nearby, along with a video identifier. At the segment interval boundary, the recorded file is hashed (SHA-256), and part of the hash broadcast alongside the next key exchange. This guards against subsequent changes to the recording file (however manipulation could occur before the recording is complete).

The video file is encrypted with AES-256, a cryptographically secure encryption function and choice of key length. This video is uploaded to be available at the URL broadcast to listening devices and deleted locally. All video processing occurs in RAM to reduce the probability of file restoration or key recovery. The Bluetooth packet structure is outlined in Figure 12.

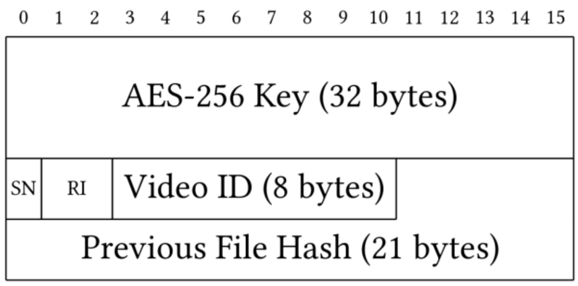


Figure 12: Key packet byte structure: AES-256 encryption key, SN - Packet sequence number, RI - Client reconnect interval, Video ID for file retrieval, first 21 bytes of last recording SHA-256 file hash.

The Camera Info Service provides meta-information, version information, a 'friendly' name and location details. The file location is stored at the URL below, with the following address structure:

<SCHEME>://<ADDRESS-OF-FILE-STORE>/<HEX-ENCODING-OF-VIDEO-ID>.{MP4|JPG}

---

[48] Raspberry PI, https://raspberrypi.org
[49] Node.JS, https://nodejs.org
[50] OpenSSL, https://openssl.org

Listening Client

Listening client runs on a user's smartphone as a passive application in the background, collecting keys from nearby cameras. These keys are stored for later use within the application. Whenever the user wants to see when they have been the subject of a recording, the discovered cameras and video segments are organized into groups based on their proximity to one another. If the user requests to play a video from within the application, the appropriate encrypted file is retrieved from the URL originally provided by the camera. This file is then decrypted locally using the AES-256 key included in the original key packet, then played back. Implementations of the client have been made available for iOS[51] and Android[52] though, due to background processing limitations on iOS, the key listener is currently less reliable on this platform.

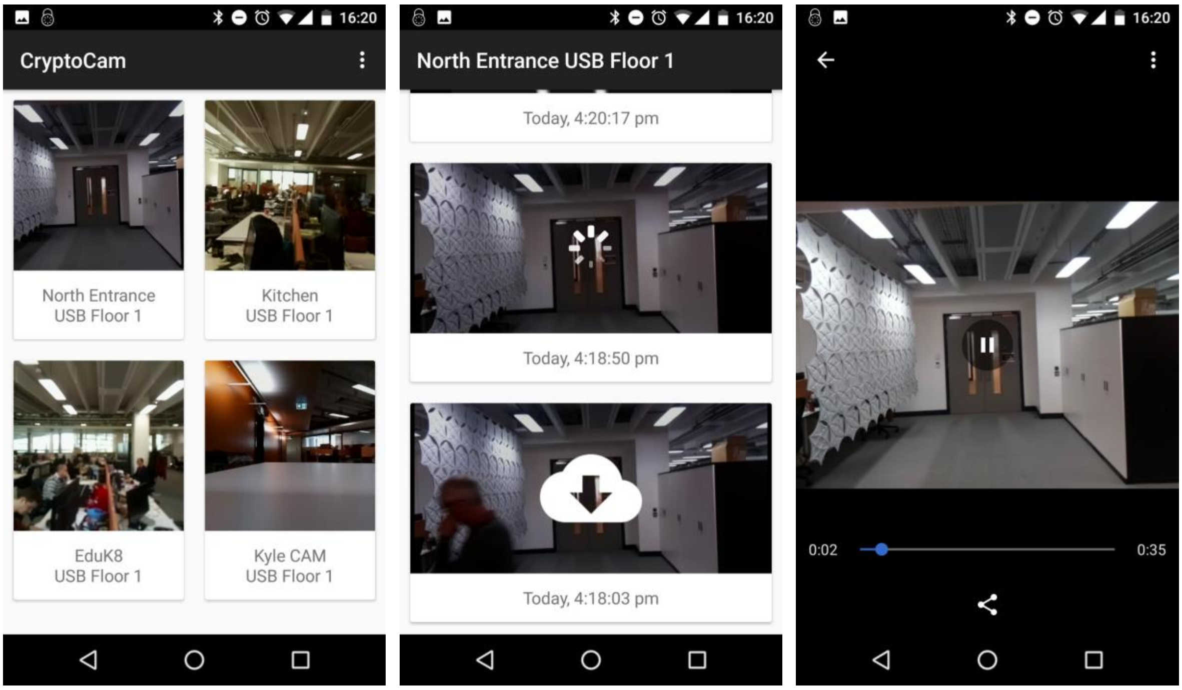


Figure 13: CryptoCam Android Application.

Finding Nearby Cameras

Using local broadcast technologies, I posit that CryptoCam can provide reliable communication, not only of its presence in an area, but also further camera meta-data. CryptoCam advertises at regular intervals with Bluetooth Low Energy broadcasts and can be interrogated by clients for further details. To address the discoverability of

[51] Apple iOS, https://apple.com/ios
[52] Android, https://android.com

cameras users need to be able to reliably and easily discover when they were/are being recorded. Each potential subject of a camera in the immediate vicinity of a camera should be notified of its presence and perhaps, depending on the application scenario, its purpose. Other meta-information about the camera, how the data is handled, length of storage, audio recording etc. are all important considerations for users near a particular camera. Using Bluetooth characteristics for static configuration variables, the camera configuration can be provided to users (outlined in Table 4).

Table 4: CryptoCam Descriptors

Characteristic	Options	UUID
Name		0001
Mode	auto, manual, delayed	0002
Location	coordinates, description	0003
URL Format		0004
Key		0011

All UUIDs are prefixed with the CryptoCam ID: cc92cc92-ca19-0000-0000-00000000####.

Simple File Access

A right of access to footage is a crucial principle of CryptoCam. Simple, reliable exercising of this right with CryptoCam is achieved through being in the vicinity (Bluetooth range) of a camera and retrieving encryption keys. This provides a user's right to access, with the scope of their access pre-determined by the camera configuration. Using the encryption keys obtained from the camera at the point of capture, users can pull the relevant recording and decrypt its contents for local playback. This process provides a balance between secure and verifiable access, and simple and readily available access. Crucially, this process is automated, requiring minimal direct administration.

Non-Exclusive Design

The features of CryptoCam outlined in this paper aim to address some of social and procedural issues around CCTV and private camera operation. However, many of these features could arguably be considered to provide unacceptable restrictions for a camera operator. The features of CryptoCam have been designed to be non-exclusive, in that the level of implementation of each feature can be interrogated by client devices and differ based on operator preference.

4.8. CryptoCam & Descriptive Data

Technological solutions such as CryptoCam provide an opportunity for increasing the wealth of information we capture about ourselves. Moving beyond often abstract metrics which we consider in isolation, I wished to explore how we can exploit this data to create a story of the user's day. The existing data/meta-data users collect could be used to index the information overload presented by fixed placed cameras. Each camera's subject and the film contents require context to understand where in the mass of collected film there are interesting points. We have already, with CryptoCam, found a potential solution to the dead film where no subjects were present, now can we index this footage further exploiting this available rich data source.

4.8.1. Privacy Implications

In providing a more open approach to data sharing we obviously introduce some privacy concerns. Camera footage is raw and largely unfiltered, motion sensors reveal presence, non-anonymous access procedures (smart card logs etc.) reveal identity. In studying a means to break down some of the procedural barriers to data access to things like CCTV cameras and moving beyond this with concepts like CryptoX, we need to consider the privacy implications. In developing the concept of CryptoCam one of the very motivations behind this was to move the ownership of data about individuals under their control, thus providing a system which is fundamentally privacy conscious – after all, who better to control their data than the user themselves.

However, conversely, open access procedures could be ripe for abuse too. Lack of control over datasets leads to potential leaks in data, with users sharing footage or using it to harass and demean others. The question underpinning this, I believe, is one of whether this should happen in public or private? Providing a more open set of access procedures means that while there are more opportunities for abuse of the data, there are also more opportunities for using the data to defend, beyond some of the more complex Object-based Media scenarios I discuss.

This project provoked a large discussion within our research group and within the University more broadly. CryptoCam did not really fit ethics rules and procedures within the University nor UK regulations. After all the ICO states that CCTV should viewing should be restricted only to those who need it, this "need" is subjective arguably, should subjects not "need" access? In gaining ethical approval for the project I had several ups and downs. We had to develop a study that contained safeguards to protect the subjects of the footage, while at the same time not compromising the very concept itself. An example of this tension (or antagonism) was in the requirement that if a participant wishes to withdraw at any time, they should have all their data deleted. CryptoCam footage never leaves the camera unencrypted, furthermore on the file stores we cannot even see which footage came from which camera. This was a design decision on our part to reduce the capability of censorship from the operator, however this was at odds with the University ethics requirements. This was further compounded by the raw nature of the collected footage, we could not control who walked in front of cameras nor adequately solicit their consent. Typically for the equivalent CCTV camera, signage and access control suffices. We were unable to offer the same assurances, the key bleed problem described above meant that, with the current iteration, we could not ensure that only subjects of the camera received access.

4.9. Proposed Study

This concept, I believe, warranted a thorough evaluation, however the issues highlighted above meant I had to compromise with the University on some aspects of the study and data access procedures. I had to maintain admin access, place signage in all camera areas and be able to delete data from specific cameras. To evaluate the concept of CryptoCam I constructed a user study to asses this challenge to existing CCTV infrastructure and perceptions. I chose a study that would not only highlight our approach to information sharing but also build a meaningful application of this footage for users. An obvious challenge to running a study with a technology such as this is that many of the scenarios and use cases we have discussed so far rely on ubiquity of CryptoCam cameras. As this was an insurmountable problem for our user study, I decided to choose a constrained environment, the workplace. I aimed to deploy cameras around Newcastle University's Urban Science Building covering a range of spaces. Meeting rooms, desk spaces, kitchens and other general break out spaces were all covered. The CryptoCams deployed were all no audio cameras following the automatic upload and share procedure outline above. Cameras were to be set up as wide angles to avoid capturing detail and provide more of a contextual shot, reducing potential privacy issues with key bleed.

In order to provide a more compelling experience to our participants, and demonstrate the possibilities for more open camera sharing policies, I was also to construct films of participants day. As part of the study I wished to request access to a range of data/meta-data about our participants. Namely their heart rate data (a fitness tracker was provided if they did not already have one), general fitness data and calendar. Deployed in a workspace I hoped to combine these data sets to produce compelling films of their day, integrating descriptive personal data into the films to enable users to reflect. Essentially, I wished to use the data to index the footage, providing insight into this potential information overload. The study was to be run over the course of a week (5 working days).

The CryptoCam application ran on participants Android phones (iOS system limitations restrict CC usage). Providing basic viewing capabilities, collecting keys from cameras the users pass throughout the day. The participants were encouraged interact with the application regularly and reflect on the privacy implications of such a system contrasted with the freedom of access. The application was slightly modified to log the keys for footage to enable the next step of the study. At the end of the week I would retrieve the logged keys from each of our participants phones along with their activity and calendar data for that week. Constructing short films of their day referencing significant portions of their day denoted by their activity or calendar data or a combination of both. Used to index against the CryptoCam footage keys we constructed a custom film of their week highlighting segments of CryptoCam film where their heart rate was elevated, they were exercising, or they had a meeting near a camera.

In post-study interviews I wished to discover what our participants perceptions of existing CCTV technology was, including their perceptions or perhaps experience of access procedures related to these cameras. Then seeking reflections on CryptoCam's more open data sharing policies, whether technologies such as this could change their perceptions around CCTV, if this would make them more comfortable with cameras being a more ubiquitous environmental sensor in workplaces, homes and other public places. Participants were to reflect on the films we had created for them, if such an application of ubiquitous cameras also influenced their perceptions around CCTV, moving it beyond simply a security/protective device towards a more personal reflective device perhaps.

4.9.1. Reaction

In developing the study I sought to test the following: Can technologies such as CryptoCam challenge existing perceptions around CCTV and ubiquitous cameras more generally? This was my question for a simple reason, for ubiquitous cameras to exist, providing a wealth of sensing and reflective material for people, we need to address the negativity around security cameras and the general perception of abuse surrounding cameras. Due to the constrained nature and high footfall of our research offices, I decided to try to run the study in our lab. I put in place comprehensive study and technical safeguards to address potential abuse of footage, signage to all entrance ways, prohibiting use of footage by Newcastle University for internal uses, removing sharing of footage from the application and coming to a set of camera placements that all members were comfortable with. Safe to say things did not entirely go to plan. The study was met with a significant amount of resistance, cameras were sabotaged, placements were disputed endlessly. All this took place before the study could even run. In total we tried to run this study three times and never succeeded.

I ran week long discussion on one of our internal anonymous discussion boards 9to5Work (Newcastle University Annonymous Discussion Forum), the topic was CryptoCam and whether we could run a study in our lab. I wished to discuss with members what their concerns were and what we could do to allay them.

4.9.2. Analysis

The discussion developed such that I decided to perform a more formal analysis of the content. Comments from the platform are anonymous and there is no way to determine if comments were from unique users or from different users, each comment (not user) was given a random pseudonym by 9to5Work. A total of 124 codings were produced with 17 codings developed from the analysis.

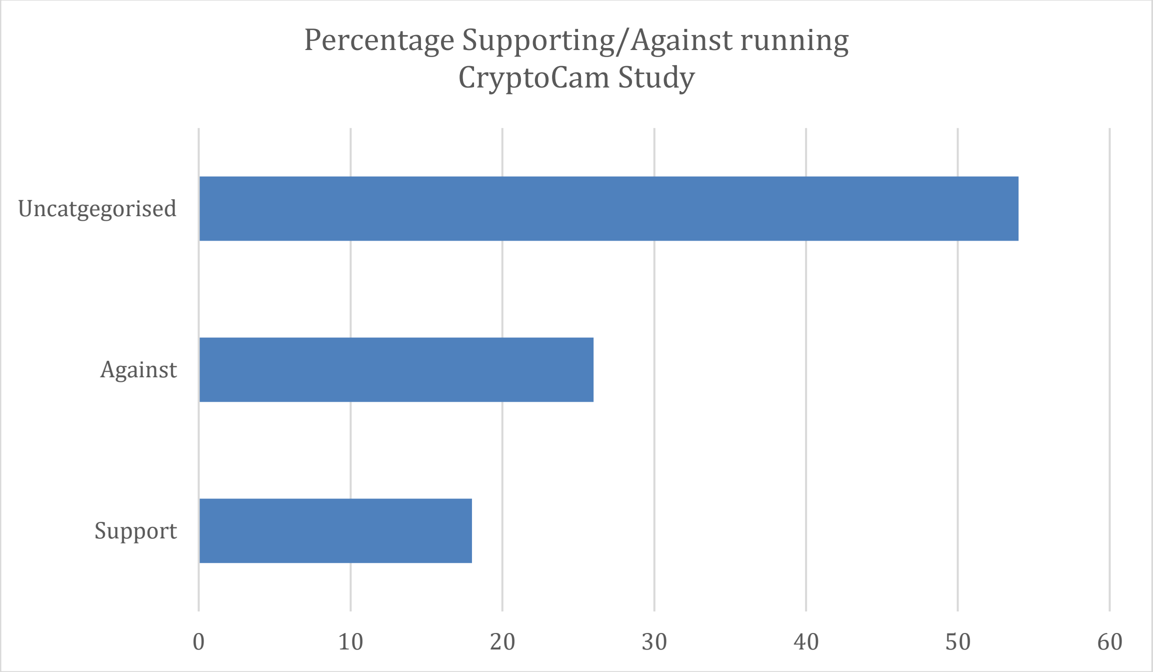


Figure 14: CryptoCam comments coded as supporting/against.

Comments were first analysed for sentiment, determining if the comment supported the running of the study (in its current form) or against running the study. Figure 14 shows that 18% of comments were in support of the CryptoCam study whilst 26% were against. As we do not have unique user information here it is difficult to gauge real statistics on overall support or lack of it within the research group as a whole. I wished to understand the rationale behind members of my research groups seeming aversion to the project. Table 5 shows the developed codings from discussion comments, each with number of occurrences.

Table 5: Codings from CryptoCam study discussion.

Code	Occurrences
Placement Issues	8
Sabotage	6
Support	18
Against	26
Study Concessions / Re-design	9
Research Un-interesting	6
Privacy Invasion	3
Data Access	5
Breach Regulations	4
Consent Issues	8
Affect Work Pattern / Environment	3
New Study Location	4
Issues with Research Team	8
Trust in Research Team	6
Ethical Concerns	7
Future Applications Concerns	3

I identified themes from coded discussion threads highlighting the following themes in Table 6. Each of the themes identified have been qualified with a selection of comments.

I have also included proposed or present mitigations, concessions and study safeguards proposed during the discussion by myself or others in an attempt to allay any concerns.

Table 6: Thematic Analysis of Anonymous Discussion.

Subject	Concerns	Mitigations / Concessions / Study Safeguards
Placement of cameras / sabotage	"but if there's one pointed anywhere near my desk I'm still going to turn it off at the wall" "that would be sabotaging another students research. Surely there are better ways to go about it?" "consent to being part of this study does this mean I can't now move freely about my workplace because of the cameras? Kitchen Teaching Room and Design Space are now off limits too?" "CCTV is not encouraged in intrusive settings; such as aimed at people's desk or social areas. It might be legal, but it's not ethical"	Each camera placement was subject to a dispute process. People who were situated permanently within a cameras line of sight could ask for it to be moved. Signage was added to all entrances with details of the project and the fact that cameras we're recording.
Consider Research Uninteresting	"This technology is in my opinion uninteresting (what's new about cameras) intrusive (as I don't want to be filmed and neither do my colleagues) and ultimately don't provide anything 'playful' or 'protective' as it was originally pitched"	Colleagues have no right to stop research because they deem it uninteresting.
Future Applications	"there is a big missing YET in this sentence (sic) the system cannot YET easily match images to recognise you BUT it doesn't mean that it wouldn't be able to do it in the (near)future"	Footage to be destroyed after study. Development of these types of invasive technologies ideated in discussion will continue irrespective of this study.
Open Data Access Systems	"Do you really want anyone to have access to camera in the workplace"	N/A
Increasing Prevalence of Cameras	"I don't like male researchers who I don't know have access to images and videos of me. When I	N/A

	have spoken to my friends that they are really sad that I am recorded at work." "cameras in the workplace are an emotive enough issue that it's unlikely you'll be able to convince everyone" "because we don't have enough surveillance and datafication in this country already..."	
Alternative Study Design	"book a large room in the building, recruit participants who want to consent, and then have them use that room as their office for the week of the study" "there is a new possibility offered of creating footage which can only be done if you hand over the decryption keys you obtained during the study" "asking those who don't want to be around to work elsewhere might be the best option for those who don't want to be recorded and keeping the study environment 'natural'"	N/A

4.9.3. Trust

One of the more significant themes of this discussion is one of trust. A number of comments referred to a lack of trust in the system doing what it was supposed to do. Several participants commented on the reliability of the encryption being used, the access levels and key bleed issues. There was also significant discussion around trust in the research team. To an outside observer and particularly a layman a camera still looks like a camera. This perception issue runs deeper than just what is happening behind the scenes. Years of examples of abuse of cameras and its raw unfiltered nature appears to have created a social impasse with people equating a camera to a person constantly watching them. Constructing technical safeguards to protect a cameras usage alone may not be sufficient to change perceptions. Furthermore, there is always an issue of trust in the team behind a product, after all a product or service's is only as secure/guaranteed as an engineering team can or choose to make it.

4.9.4. Further Reflections

Alas, I was unable to convince a noisy minority, their obvious and ingrained animosity towards CCTV, with its many examples of abuse by governments and companies. Many were simply unwilling to try something that could remove some of the possibilities of misuse that are possible with CCTV. Genuine concerns were conflated with contrived abuses and use cases that were often not only beyond the scope of the project but beyond even the largest consumer technology companies in terms of machine intelligence.

There are several aspects which I will happily reflect upon here where our approach to this study were flawed. Cameras in office workspaces are uncommon, and becoming less common. Introducing cameras to a space where there were none before will naturally give rise to the question, well we didn't need them before so why now?

Our lab has a significant civic engagement, bordering on activist agenda. Often provocative works are used to generate debate around particular topics, being a computing lab, we as a group are more aware than the average group of the potential for technology. This, I feel, manifested itself in a few ways in this research project. First people did not trust that these cameras did only what we said they did. They did not trust there was no raw admin access, a question often asked of me. Some even questioned the potential failure of the encryption. Secondly, considering not so distant future applications where we could have reliable facial recognition enabling tracking of people and deep recognition of contextual information from the film, many individuals were uncomfortable at the prospect of more cameras in their environments. Much like some of the technologies I discuss in this thesis, greater inference from datasets for personal use will inevitably have monetary value to advertisers and other surveillance uses. Combined these concerns lead to an incredibly hostile environment for deployment of this technology.

In more classical literature Goold *et al.* [30] refer to the most significant perception problem CCTV faces is the "unobservable observer". The operator with little oversight, breeds a mistrust of CCTV, despite its well documented role in criminal and abuse cases worldwide. However, reflecting on the situation I was presented with above, anecdotal though it is, perhaps this is no longer the main factor behind these concerns. Large scale data collection by governments and private organisations around the world for the purposes of monetisation appears to have given rise to an arguably more serious concern of data protection. I reflect on Möllers & Hälterlein [62] and Moncrieff *et al.* [63]

assertions that "privacy is an optimisation problem", optimisation, as they discuss it, involves revealing as little information as possible whilst also providing the required functionality. Removal of faces (blurring or blanking), lowering the resolution, removing colour channels etc. are all techniques designed to reduce the information relayed by the footage, the degrees to which this occur, the authors argue, are dynamic – based upon the context in which the camera is deployed. CryptoCam introduces a new means for dynamism, access to footage based upon locality. This can be considered a further optimisation, only those who were near the camera are provided with access. Such technologies explicitly prohibit some of the abuses (or perceived abuses) that we may attribute to large multi-nationals and governments with data. This could be the optimisation that matters most in today's society.

4.10. CryptoX

CryptoCam provides a reliable and feasible means for sharing footage from cameras to those nearby. The system architecture enables low power communication of access tokens for later data retrieval. This process could be considered lazy in the sense that the data retrieval is only done when requested. This structure for data sharing is immediately extensible due to its nature, the size of the data cache to be shared is irrespective of the distributed key size, the data itself can be anything. This led me to develop the concept of CryptoX, deploying the CryptoCam architecture beyond footage but for any data. For example, sensor data collected by your company about the building, or your local council about your community spaces. In some cases this data may be publicly available, though without adequate contextualisation or even discovery mechanisms this data would most likely appear arbitrary to many. Applying the CryptoCam principles of discoverability, access and inherently contextualised through locality, we could break new ground in getting communities interested in local data. This could be from an environmentalist standpoint, privacy or simple curiosity. Recent work in citizen data sensemaking has looked at techniques to encourage community interest in data "social sensemaking" as they term it [79], part of a larger project about public sensors from Newcastle University's Urban Observatory. Research like this demonstrates a user willingness to share, and some of the exciting prospects for sharing such information. Combined with more contextualised and situated data capture and sharing solutions like CryptoX we could provide a compelling framework for the large-scale sharing of data which may be relevant/of interest to others.

4.11. Contact Tracing for the Novel Coronavirus (COVID-19)

The key sharing protocol which underpins CryptoCam is now a fundamental of contact tracing systems that are not being widely adopted by govenments and countries all over the world. Sharing unique but non-deterministic identifiers (keys) via Bluetooth to other smartphones nearby is undergoing trials to determine its effectiveness at providing a mechanism to trace contacts of indivuduals who have become infected by highly infectious deseases such as COVID-19. This concept is analogous to some of the discussion of the possibilities for CryptoX above. The contribution of this work in providing a system to enable a coordinated but privacy centric contact tracing system where subjects determine access to their traced contact history, is one which I am very proud to say I worked on and developed with my colleagues at Newcastle University.

4.12. Opportunities for Passive Reflection

Collected personal data is often a passive process, users often have to do very little to collect this information, beyond perhaps putting on a smartwatch or fitness tracker in the morning. These kinds of passive collection activities are largely why such utilities are successful. Avoiding the need for direct user input but providing compelling data points for later reflection is an important selling point of activity trackers. The ability to go back and look at how a particular moment in my day or week affected my vitals is powerful and provides context to certain situations. Opportunities for passive reflection with film are more limited, typically any film you have captured required an active process on your part. This passiveness is however important, reflecting on something later could be argued to be more useful if you were not aware beforehand such a moment was worthy of capture. How can we make film a more passive data capture process?

Technologies such as CryptoCam offer some compelling solutions to this issue of passivity in capture. Film from cameras nearby are made available to a user with only their smartphone in their pocket. Furthermore, this introduces a passivity to sharing with friends, when taking a photo or video with friends around this can be shared with little or no input from the person holding the camera. In such an environment we could open up new possibilities for this process of passive reflection. Where did I leave this item? What was I doing at this time? Providing localised contextual information through captured film we can enable passive reflection with film.

4.13. Integration at the Expense of Privacy?

Integration of these datasets as I argue for at the outset of this chapter, introduces further privacy complications. Information that is not revealed in siloed datasets could be unveiled through integration. Indeed, this is the very reason we advocate for integration. As a working example, Google tracks a wealth of information about its users, their search queries, their location data, and general device usage information. Recent news articles would appear to suggest that Google are also interested in harvesting medical information from the NHS and other health care providers [26,37,38]. This health data is anonymised by the provider but could be rather trivially de-anonymised by a corporation like Google. Through integration of personal search terms and location data, cross-referenced with the patient treatment dates and hospital locations. This problem was broached in one of our lab projects AppMovement [28], site usage logs provided a wealth of information on their users. This information by itself revealed little more than what users were making what requests. However, when combined with the location services data scraped when placing new sites, or adding reviews the author was able to identify not only where users most likely lived but also their working patterns, workplace and more. When designing open access data solutions like those advocated by Puussaar *et al.* [80] we should be careful to ensure there are technical safeguards in place to ensure that data leaks through integration are limited.

Removing personal data from one dataset could be restored by another, for example removing all faces with Computer Vision in CryptoCam footage, similarly this could be restored with location data from another dataset.

4.14. Summary

Integration of services and datasets undoubtedly leads to greater insights and in the case of media consumption more immersive data foundations for experiences and feedback loop for underlying Object-based logic models. The ever-increasing ubiquity of sensing in our environments, both personal and infrastructural, provides a bedrock for construction of media experiences to exploit these advances.

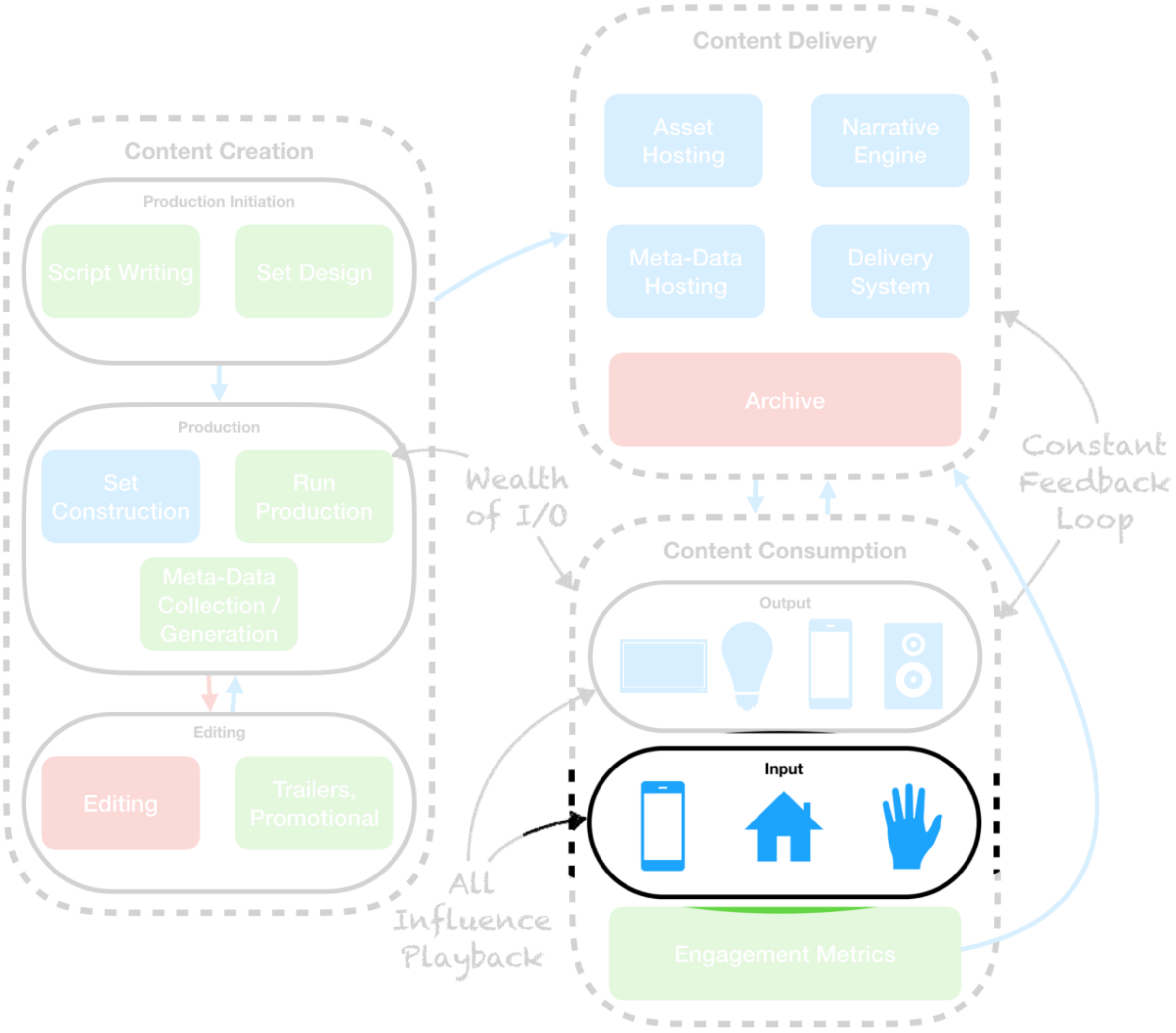
In this chapter, I sought to investigate an increased depth of understanding through integration of existing datasets. In the previous chapter Media of Things, I looked at explicit instrumentation and generation of previously non-existent data from film. However, it is incumbent upon us that we should look to exploit existing datasets. Integration of these datasets into production and consumption workflows could prove compelling additions, or even replacements, to some of these more explicit sensing technologies proposed earlier. Industry have sought to increase the integration of consumer data with platforms to sit above the existing disparate and piecemeal nature of services and devices. Following industry lead, can we provide more personally powerful data inferences providing more compelling media consumption experiences. CryptoCam demonstrates an example of this, working from a starting point of providing a previously infeasibly accessible dataset and using data to index film. I have constructed a system to enable more contextualised reflection upon this content, providing previously unavailable spontaneity to video capture. Furthermore, this concept is immediately extensible beyond fixed cameras with smartphone film captured by others and other mobile cameras coming into the fold with software deployments of CryptoCam.

More closely aligned with traditional broadcast media, I have discussed data in current productions. In Media of Things we advocate for the content delivery to perhaps include some almost raw data metrics, enabling some of the more compelling Object-based Broadcasting experiences to be realised. Similarly, with existing metrics we would argue for their separation from the broadcast media (e.g. sports metrics, vote counts etc.), thereby enabling their redeployment for other sub-systems within the consumption experience. I would advocate for a broader goal within media production for the separation of content and metrics, only with such an emphasis do I believe we can produce compelling Object-based Experiences. Indeed I am not alone in this thinking [54].

Finally, I have discussed in depth the subsequent privacy implications of some of our work and some of the societal challenges that come with deeper integration of content and services. Technologies like CryptoCam were developed with the aim of subverting, to some extent, the status quo. Moving control of our data from centralised proprietary siloes towards a move user centric approach. However, in the case of video more needs to be done to address the legitimate concerns around access and the implications of ever-increasing machine intelligence. The technical safeguards present in systems like CryptoCam do not appear to be enough and monetisation of collection of keys, for example, could effectively bypass this. Broadcasters need to tread carefully when considering how their experiences exploit and manipulate environments. Privacy is a hot topic of conversation in modern society, it should be treated with the focus and care that it deserves.

Maintaining Control

Chapter Five



Open Lab at Newcastle University

Chapter 5. Maintaining Control

In this chapter, I explore the emerging issues with the wider deployment and adoption of Object-based Broadcasting consumption experiences. I posit that increasing immersion and moving experiences from the constrained space of a single media playback device, towards an adaptive experience which exploits all available playback mediums, could lead to a loss of control. Simplistic control interfaces such as a TV remote are no longer adequate for maintaining control over the experience.

It is, however, incumbent upon us to make the best of the devices and smart technologies available in the typical home at present. There is a vast array of devices in the home that are directly or indirectly instrumented. Take a television remote for example, its power button has one purpose to toggle the power state of the TV. However can we repurpose this for controlling other devices. In doing so we can increase the interconnectivity of existing technology in the home in order to glean new information and new understanding of the current state of the home.

5.1. The Problem

Object-based Broadcasting provides new opportunities for immersive consumption experiences. As an inherently adaptive media playback paradigm, OBB provides an opportunity for increased levels of consumer immersion. Exploiting the potential wealth of playback and control devices in the modern home the concept of Object-based Media can be extended to incorporate most devices. Moving away from media as a simplistic pre-packaged consumption experience tailored to one or several output mediums, instead we can view media as a non-prepacked asset. In each environment, the way the media is presented could be different, making the best use of the available hardware. In such experiences, we could breach new levels of user immersion in content both in entertainment context and informational.

There is however a natural paradox with ever increasing levels of immersion, maintaining control. Existing media consumption experiences, while simplistic in nature when compared to the potential for Object-based experiences, are rather simple to control and navigate. The linearity of media sees to that, four controls can govern the entire experience, play/pause/fast-forward/rewind. This age-old interface to media becomes less and less appropriate as media becomes more immersive and includes many devices as potential input and output mediums.

5.2. Problematic for Control Interfaces

Object-based Broadcasting as a concept is not intrinsically linked to any one playback environment or set of devices. In its truest sense, the concept can be incredibly flexible making best use of available screens and input/output devices. Any number of playback mediums could be influenced by the media, any number of control interfaces could be providing the required control. Existing control interfaces to devices will no longer suffice, play/pause fast-forward and rewind cannot cover the degrees of freedom offered by an immersive playback experience. The movement of a particular utensil in the kitchen could move the playback to a new location, or a choice is made through voice. I need to investigate control techniques which provide the required degrees of freedom for control with Object-based Media experiences. Smart Home device control, I feel is emblematic of this. As traditionally 'dumb' devices are networked this opens up new possibilities for communications, but crucially these new devices offer a wider range of characteristics for control to the user's: smart lights offer colour; smart music systems offer direct access to a catalogue of songs; and smart appliances offer dynamic feedback and more intelligent programs based on user contextual information. These advances in technology typically came with corresponding smartphone applications or voice interfaces to provide the required control. Gestural control, furthermore, is becoming more and more prevalent as a control paradigm. Providing more expressive and 'natural' interfaces to devices. In this section, I will explore how we can similarly exploit these works on device control to provide compelling control interfaces to Object-based experiences.

5.3. Related Works

There is a rich history of research in gestural interactions as a means of machine control in the home. Many have relied on instrumentation of the space; for instance Put That There [12] and latterly g-stalt [106] – an implementation of the gestural interactions demonstrated in the film Minority Report (2002) using Vicon motion capture system. However, these large-scale installations of infrastructure are in tension with the typically ad hoc nature of technology change in the home. Assemblies of more pragmatically installed sensors (e.g. Kinect [73]) offer a more sympathetic path for adoption, yet they often fail to deliver the promised ubiquity of action – confining their stage to the immediate space in front of the device. Rather than instrumenting the space, a common strategy is to instrument the individual devices for expressive gestural input. FingerSense [95] is a camera based prototype that tracks and identifies fingers interacting with a "button" using colour tags. SideSight [13] allows multi-touch interaction around the edges of small devices on adjacent surfaces using IR light. Project Soli [96] uses high-frequency short-range radar sensing to track nuanced finger gestures performed in front of the sensor. While allowing a gradual adoption, the promise of a universal (at least within the walls of the home) gestural language becomes diluted over multiple manufactures and sensing technologies. Alanwar *et al.* proposed SeleCon [2] a pointing-based control system for IoT device control. The authors explore using a wrist-worn ultra-wideband sensor to detect which device a user is pointing at and gauge distance to said device. Alanwar *et al.* developed a set of interactions to control these devices. However, this system requires modification of all devices in the home and wrist-worn sensing hardware not present in any commodity hardware.

An alternative to embedded sensing is to use a proxy that then communicates with the device: the remote-control model. The IKEA Motion Controller[53] is a recent exemplar of this; in this case gesturing with an accelerometer equipped puck. However, in providing limited gestural control alongside more flexible device inclusion, a puck is required for each control scenario and the interaction control is pre-configured rather than defined at point of interaction.

Discoverable and contextually aware interfaces to devices have been explored with smartphones as a proxy device. Laput *et al.* and Xiao *et al.* [53,102] exploited an electromagnetic field sensor on a range of devices to identify electronic devices unique EM signature. Xiao used this to present the relevant interface to a particular IoT device when placed near to the device. Laput looked at a range of application scenarios including authentication of a user entering a room. Google's Physical Web[54] uses Eddystone-URL beacons to advertise a particular device's web-based interface. These works demonstrate a research and industry interest in development of discoverable and

[53] IKEA Tradfri Motion Controller, https://www.ikea.com/gb/en/products/lighting/smart-lighting/
[54] The Physical Web, https://google.github.io/physical-web/

immediate interfaces to devices and crucially providing context in IoT scenarios. Establishing this additional context during interaction opens up opportunities to provide more subtle and expressive interfaces. Jentz *et al.* [47] produced a 'wand' trackable by an Augmented Reality headset for interaction with devices. Apple Pencil[55] provides a more expressive input to touch-based devices. In some of my previous work [100], I explored using a wrist-worn IMU to add expressiveness to touch, some of the control demonstrations shown by the authors are analogous to IoT device control, e.g. post-interaction fine scrolling of text on a touch screen which provided deeper and fine grained control of the scroll is comparable to continuous control of the brightness of a bulb after switching it on.

5.3.1. Complementary Control

I explore the opportunity for a wrist-worn IMU to provide an alternative means of control to complement existing IoT device interactions. At the nexus of existing interfaces is where I see the potential for exploiting gestural control, including otherwise "dumb" devices in interactions and combining the functionality of smart devices to develop new interactions which provide adequate coverage of immediacy, context and control. This can be used across an ad-hoc collection of devices and manufacturers. It exploits the emergence of APIs for these devices (Apple HomeKit, etc.) that are exposed through a common communication medium (the home WiFi network and bridges to Zigbee devices, etc.). Currently interoperability is achieved through the efforts of Open Source projects; we might be hopeful that these proprietary protocols become open in time. The prevalence of wrist worn smart devices with accelerometers provides us with a convenient remote proxy for interaction that I co-opt in this work. This proxy is in Heidegger's [36] terms literally *present-at-hand* (our focus is the interaction rather than the device) unlike the traditional *ready-at-hand* easily-mislaid remote control. In addition, gestural control offers a site of explicit action and feedback; and a means of identifying an individual user. In addition, I believe there is an opportunity to leverage the sensing hardware in wrist-worn smart devices in conjunction with underexploited IoT device information, e.g. low-level network information and device status, to provide a platform for seamless control of IoT devices. Regardless of the sensing technology implemented – designs for in air gestural

[55] Apple Pencil, https://apple.com/apple-pencil

interactions, for multiple possible targets, must address some common hurdles, previously identified in the literature [2,4,59]. Specifically, that of device identification and gestural (dis)connect. Often this is a matter of Proxemics [33]; where the gesture occurs in contact or in close proximity to the target. However, this is not always possible; where the target is out of reach we typically defer to a pointing strategy [12]. Similarly, Attentive UI [94] builds on social queues, such as eye contact, to disambiguate targets. Once the target is engaged there is then the related problem of disengagement; the Midas Touch [46].

5.4. Expression in Control

In approaching this problem I looked at the evolution of other control interfaces for inspiration. Interfaces to technology have evolved as the capabilities of their controlled devices have increased. Mobile phones of 15 years ago are almost unrecognisable in comparison to today's devices. As the functionality of the devices increased with increased mobile processing power and display technology so too did the control interfaces: complex touch interactions replaced simplistic directional input; and quick type keypads replaced with full touch keyboards. Expression in control with pressure sensitive screens and smart styluses like Apple Pencil[56] provide compelling control paradigms for these devices. Similarly, light switches are evolving to meet the new characteristics available with smart light bulbs, colour options, colour temperature etc. each require new control interfaces to function. TV remote controls are also evolving to new more interactive content, touch pads and motion controls are providing new compelling interfaces to media and games on TV sets. PC's, tablets and games consoles continually expand their dynamism of control through body tracking, motion control and more.

In each of these examples above, we see that with industry advances in technology we have seen control interfaces similarly innovate and expand in their degrees of freedom of control. I view this as a comprehensive expansion of the expressiveness of these interfaces as a whole. Industry and researchers [100,103] appear to agree that enabling

[56] Apple Pencil, https://apple.com/apple-pencil

users to be more expressive in control is the solution to controlling complex systems and parameters.

Object-based consumption scenarios greatly increase the potential complexity involved with controlling the media playback. Simplistic controls no longer suffice. How can we construct a control system that maintains users control over the media playback in a compelling and adequate manner? To understand this problem and approach a solution I worked on a compound control interface to devices in the home, including the TV.

5.5. Expressive Things

The pervasive interconnectivity of devices, sensors and services offered by the Internet of Things (IoT) is set to revolutionize our interactions with both the environments in which we live, and the people with whom we share them. One environment that IoT is expected to have a particularly significant impact upon is the home. We are already seeing the development of a multitude of Smart Home devices that exploit new opportunities, e.g. automation and context awareness, to enrich aspects of the home including lighting (Philips Hue[57]), temperature (Nest Thermostat[58]) and entertainment (Sonos[59]). The popularity of these devices is predicted to rise significantly in the coming years, with two thirds of consumers planning to purchase an in-home IoT device by 2019 [1].

IoT devices in the Smart Home increasingly offer a largely unprecedented level of convenience at an affordable price. The increase in convenience comes through greater interconnectivity of devices and bandwidth for control with new parameters such as lighting colour and timed control. I argue that three important factors of this convenience are:

- Immediacy (the speed with which an action can be conducted)
- Context (how much information about the intended interaction can be inferred from contextual factors)

---

[57] Philips Hue, https://www2.meethue.com
[58] Nest Learning Thermostat, https://nest.com/thermostat
[59] Sonos, https://sonos.com

- Control (the depth and ease of characteristic control). I demonstrate the importance of these factors situated in existing IoT device control

Many visions of the smart home [35] imagine that our interactions with devices will be minimal and passive, drawing on sensor data in combination with artificial intelligence techniques to limit the input required by the user. Yet, emergent smart home application scenarios show us that the direct expression of user intent will still be an important aspect of interaction. The need to immediately change or express preferences for automated configuration of lighting parameters, temperature conditions and audio entertainment, should be addressed by interfaces that allow users to adequately express their intentions. Currently, these devices are primarily controlled via an array of smartphone apps, and their functions are increasingly being coordinated via a platform such as the Apple HomeKit[60] or integrated into speech interfaces such as Amazon's Alexa[61] or Google's Home[62]. The piecemeal nature of the home lends itself to such a structure, as devices and interfaces of many manufactures are accumulated over the lifetime of an individual – as such, these frameworks are developed to cope. However, whilst these platforms unify the interfaces for control, smartphone-based interaction can lack immediacy and convenience and voice interaction is not well suited for continuous and precise control.

I look to exploit the most favourable parts of each interface to IoT devices. I reason that while any one interface does not adequately cover Immediacy, Context and Control a combination of such interfaces has the potential to do so.

5.6. The State of the Art

The motivation behind this work, is to explore how gestural interaction could complement existing IoT device interactions. Our user study, and subsequent interviews, exposed the gaps not only in existing interfaces, but also the limitations of using gestural control in isolation. The interaction model we develop further highlights the potential for over-complication of control which was traditionally simple and accessible. Working at the nexus of existing interfaces to IoT devices, we look to enable

---

[60] Apple HomeKit, https://apple.com/ios/homekit
[61] Amazon Alexa, https://amazon.com/alexa
[62] Google Home, https://home.google.com

cross-interface interaction, with gestural control as the "glue" to bring these interfaces together. I propose that issues with existing IoT interfaces follow three main themes:

1) ***Immediacy issues***, the interaction is slow or requires multiple steps to complete (e.g. walking to a physical control or picking up a phone then starting an app).
2) ***Contextual issues***, where the interaction does not inherently provide adequate context information to avoid potential mis-interaction (we cannot know where exactly a smartphone issuing a command is, thus context information must be explicitly selected).
3) ***Control issues***, the interaction does not provide adequate rich control (e.g. selecting between intensity, colour and mode) or continuous control (e.g. level of brightness or volume).

Table 7. Comparison of existing interfaces' Context, Immediacy and Control

	Context	**Immediacy**	**Control**
Physical Interface	Yes	No	Simplistic & Pre-Defined
Smartphone	No	No	Complete
Voice (Alexa/Google Home)	Limited	Yes	Simplistic
Gestural (ET/Kinect)	No	Yes	Complete

Table 7 provides an overview of some of the more common consumer and research interfaces to IoT devices. Each interface has its strengths and weaknesses in Context, Immediacy and Control, with no interface covering all three adequately.

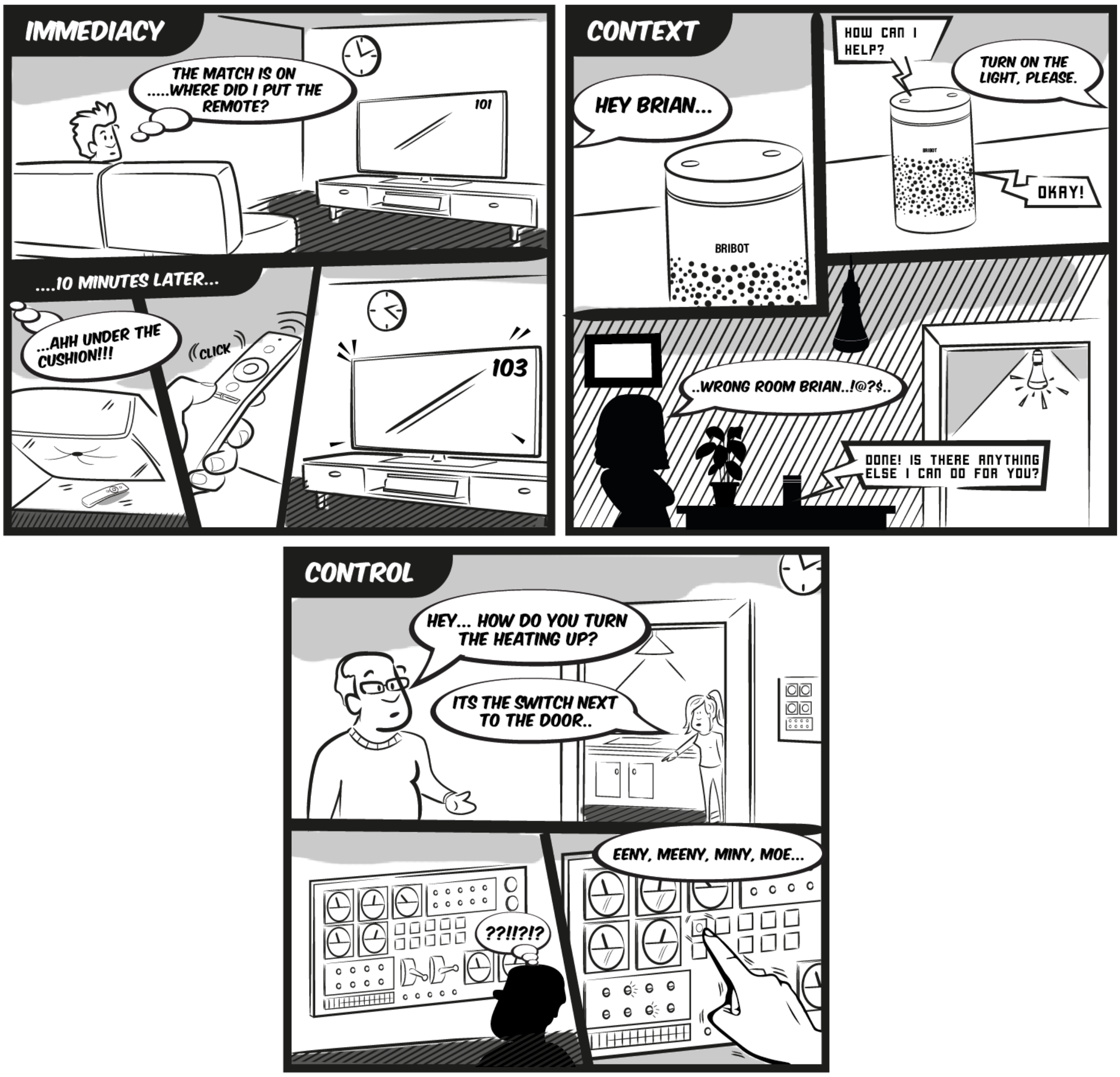


Figure 15: Cartoon strip highlighting the three highlighted issues of Immediacy, Context and Control.

Figure 15 illustrates a common example of the failings in Immediacy, Context and Control for typical interfaces to smart home devices. In the example for immediacy the user is unable to find the remote in a timely manner. The voice assistant turning the light on in an unintended room illustrates a common contextual issue. Finally, we have an example of a lack of simplicity in control – such an interface provides maximum control at the expense of usability, balance is key here.

Exploiting increased interconnectivity of devices and services, I explore the potential of using gestural interactions performed with a wrist worn inertial measurement unit (IMU) complimenting and exploiting existing control interfaces. I believe that this approach can offer a number of advantages. Basing interactions upon body gestures means that parameter changes can be quickly initiated and performed by the user without the overhead of opening a mobile application or locating a wall-mounted control interface, for example, thus providing immediacy. Equally, such interaction has the potential to offer rich and expressive control over devices because it can exploit the precise and complex gestures we can make with our hands and arms, in particular in the context of continuous parameters [100], providing rich control. Existing interfaces to devices can also be included in such a system, a light switch can provide the context of the device to control for instance, or initiate the interaction crucially without modification. Finally, tracking gestures with a sensor found in the smart watches and fitness monitoring devices that many people already wear on their wrists [29] means that the approach can be implemented without the need for additional expensive infrastructure. Enabling a software solution with readily available hardware also provides the opportunity for continuous improvement and expansion of control as Machine Learning techniques on the signal data improve.

This research lays the foundations for the design and development of Smart Home interfaces based on this approach, by presenting the following contributions:

1) Interaction elicitation and interaction-preference studies that reveal possible cross-interface interactions for controlling Smart Home devices
2) A conceptual model that aims to guide design by characterizing the outcomes of these studies in terms of: Thing Identification, Characteristic Selection, Control & Disconnection
3) A proof of concept technical implementation that demonstrates how the form of interaction envisaged can be realized across multiple application scenarios using currently available hardware
4) A set of user application scenarios that demonstrate the power of gestural control for the Internet of Things and novel media constumption experiences (Object-based Broadcasting)

5.7. Study

We cannot simply introduce a control paradigm with more degrees of freedom without eliciting what type of interactions will work. In order to achieve this, I conducted two studies to elicit interactions to control Smart Home devices. My aim was to determine a set of appropriate interactions for devices in the home when considering cross-interface interaction.

Throughout this chapter, I discuss the potential for loss of control with Object-based Media. Dedicating an entire chapter to discussing this problem, and working to develop a solution. In designing the Expressive Things study, and the concept more generally, I viewed the expansion of controlable parameters as an opportunity rather than a challenge. Providing participants with the possibility to include multiple interfaces in their interactions, constructing convenient and expressive control systems. In doing so, I hoped to generate a set of interactions which had higher degrees of freedom for control. Application of these interactions were purposefully broad, ensuring that generic control interfaces were produced, such that they were appropriate for media control, Smart Home control and more.

5.7.1. Interaction Elicitation

Gestural control with a wrist-worn IMU and mid-air interactions trackable from such a sensor is a largely unbounded and broad space. Inclusion of existing interfaces in the interaction and multi-interface interactions further compound the problem. To adequately understand and classify this interaction space I conducted an elicitation study. Elicitation of such interactions in a novel trackable medium follows previous work [65,101] allowing users to develop interactions to match devices and their characteristics. To explore the possibilities for wrist-worn sensing interactions with IoT devices we conducted an interaction elicitation study followed by an interaction-preference study. To determine preferred interactions for each control scenario I then showed the most popular developed interactions to a new, larger set of participants. At the end of the study I aimed to have a set of user designed interactions ranked in order of preference for each control scenario.

To elicit interactions from our participants we followed the structure of Morris *et al.* [65] & Wobbrock *et al.* [101] interaction elicitation studies for Kinect and table top interactions. Users were free to develop interactions that they deemed trackable by the wrist-worn sensor in conjunction with existing interfaces. The purpose of this study was to elicit a range of potential interactions within this space.

We recruited 10 participants average age 27 (SD=5.2). To assist participants understanding of what can or cannot be tracked by a wrist-worn sensor they were asked to wear an OpenMovement WAX9 IMU [66] on their dominant wrist. At the beginning of the study they were shown a representation of the device tracking their wrist movements as a 3D model, as well as some live statistics (force, roll & pitch). This was to try and focus participants on what could be tracked by the sensor, they were encouraged to discover what can and can't be tracked at this stage. This visualization was deactivated before the participants began designing their interactions to reduce any bias introduced by the representation of the data.

Some examples of the devices under control were lighting (e.g. brightness), TV and Music (e.g. volume, play/pause), temp (e.g. on/off, away**.** The coordinator demonstrated a device characteristic, and then asked the participant to invent an interaction to control it. This was repeated for each of the device characteristics. Video footage was captured of each interaction invented by participant and feedback to their interaction was provided through Wizard of Oz techniques. The participants were asked to think aloud during the performance to enable the coordinator to provide suitable feedback. Participants were prompted by the coordinator to come up with multiple interactions to control each characteristic, to reduce any bias implied by the command, coordinator or environment [101].

To reduce the complexity of interaction development for participants we let them assume that the device is already selected, and they could largely ignore accidental interaction with other devices. This decision may seem to remove an obvious problem with any such control system. However, an initial trial run of the study found that including device selection was largely redundant, as most participants were selecting with voice or pointing at the device for single selection. I felt that device selection was a solvable problem (as demonstrated by my prototype implementation later in this chapter) and that providing participants with greater freedom to explore beyond the selection issue was acceptable and would still provide me with the desired output of a set of cross-interface interactions.

Findings

Overall participants produced 244 interactions with, on average, two interactions produced per command, per user. Following on from the study design by Wobbrock [101] and Morris [65], I first coded each participant's interactions. Interactions were coded detailing how the characteristic was controlled, and the degrees of freedom utilized. The coding was carried out through analysis of video and audio recorded of each participant session.

I coded our interactions as follows, arm position (outstretched, raised up, down by side and toward device), interaction mode (discrete or continuous), degree of freedom used (roll, pitch, yaw, force, compound) and the interaction developed (i.e. Figure 16). Taking the coded interactions, I performed thematic analysis on the dataset grouping similar interactions and degrees of freedom used. I highlighted themes first by considering interactions that provided discrete and continuous control. Then further sub-divided into similar interactions, for example swiping left/right, swiping up/down, and continuous roll of wrist.

Participants, having been shown a representation of the sensor data at the beginning of the study, demonstrated a good coverage of each degree of freedom throughout their sessions. Roll, pitch, yaw, force and a compound of degrees were each used at least once by each participant on average. Participants were also largely consistent in their use of discrete (e.g. swipes or taps) or continuous interactions (e.g. wrist roll or sweeping motion up and down) for control over characteristics. State changes were exclusively controlled with discrete interactions and value changes were predominately controlled by continuous interactions. Compound control (e.g. twirl in the air), the most prevalent interaction technique, combined several degrees of freedom to produce a single discrete interaction. Participants often commented throughout *"I want to copy what's on-screen (TV)"*, *"like a brightness knob"*, as such roll and pitch were more popular for brightness and volume, pitch and yaw likewise for channel changing, fast-forward and rewind.

As expected, lots of interactions could potentially overlap with each other. We asked participants to consider the most appropriate action, however they were allowed to consider the device already selected. Disambiguation of interactions for devices is therefore a vital component of any expressive interaction system for IoT device control.

Again, disconnection from a device or set of devices once control is complete is an important facet of any gestural control system. Participants were consistent, dropping of the wrist, down by their side or significant pauses in the interaction would disengage the system. In cases where an explicit final selection was required participants performed forceful swipe or tap interactions that resulted in selection then disconnection.

5.7.2. Interaction-Preference Elicitation

Once themes had been highlighted from the interaction elicitation for each type of characteristic control, I highlighted up to five of the most popular interactions for each command. I produced new videos of each interaction with subtitles explaining each stage and detailing the output. I then recruited a further 20 participants average age 30 (SD=4.1). Participants were given the same command description as participants from the first study then were shown up to five videos of previously invented interactions for the command, participants were encouraged to mimic the recorded interaction to make more informed assessments. They were asked to rank each interaction in order of preference. Commands were presented to each participant in a random order to reduce any inadvertent comparisons they may have made between command interactions. Semi-structured interviews were conducted with each participant when they had concluded their rankings. The following questions guided the interview:

- Is this type of control something that you would like to use in your home?
- What would prevent you from using this type of control?
- Do you find the interactions comfortable and appropriate, would you do this in front of others?
- Do you think a system such as this would give you a more compelling reason to own, wear and charge a smartwatch?

Results

To summarize the preferences of the respondents towards different interactions, the Plackett-Luce model is used [15,32]. The Plackett-Luce model treats the whole set of preferences of each respondent as one observation, and assumes there is a strength associated with each interaction. For each question, each interaction has a strength (0-

1), and the strengths of all interactions sum to 1. The higher the strength, the more likely an interaction is to be ranked higher.

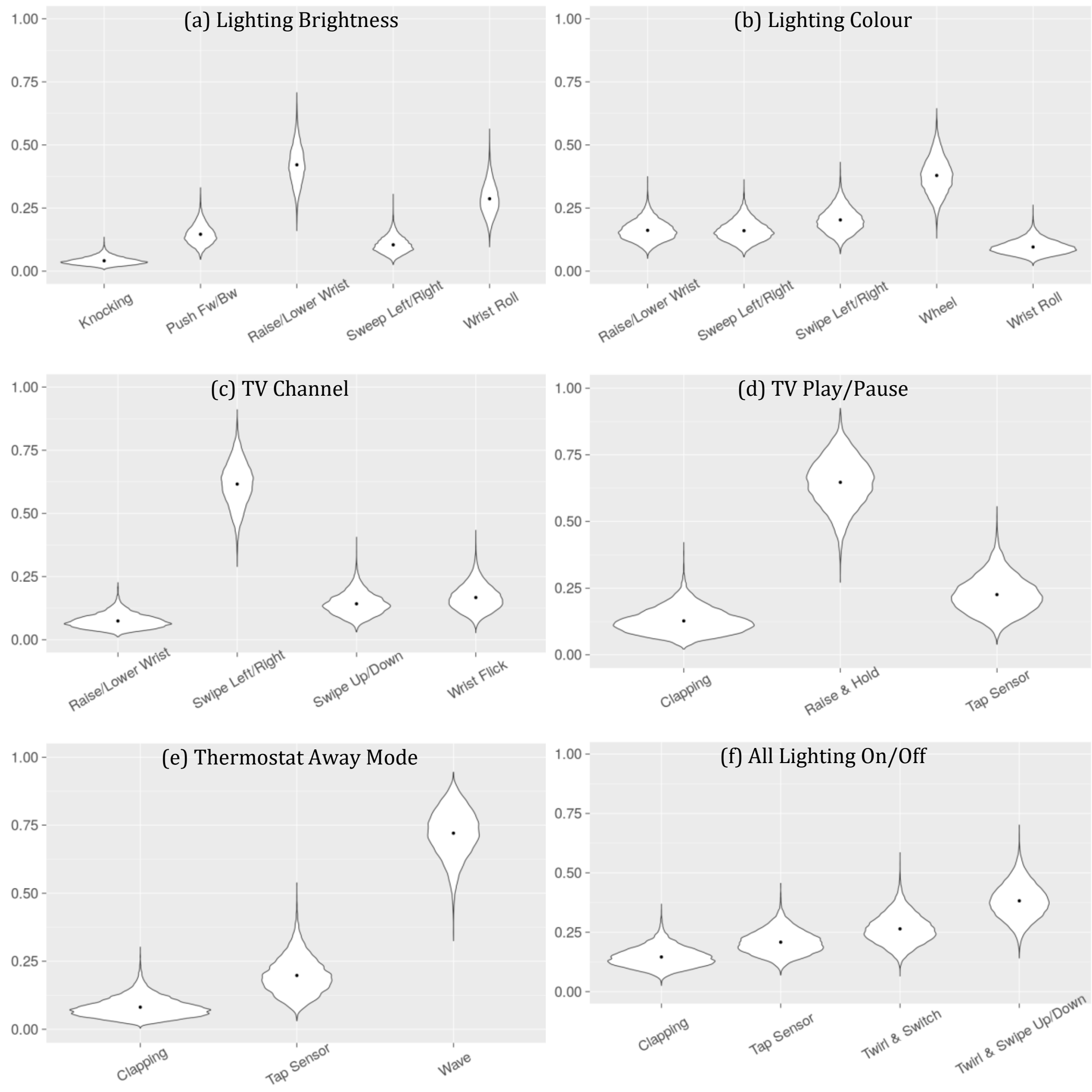


Figure 16: Violin plot of strength against interaction. The dot in the middle represents the mean strength. Higher means higher ranked, shorter means lower uncertainty.

Figure 16 (a) illustrates the violin plot of the strength of five different interactions for lighting brightness. Raising or lowering wrist has the highest mean strength and therefore is the most preferred interaction, albeit with high uncertainty (height of the "violin"). On the other hand, knocking is the least preferred interaction with low uncertainty. Figure 16 (b) shows the wheel gesture (cycling through colours with a mid-air wheel motion) to be the preferred interaction. Figure 16 (c) shows swiping left (change down) / right (change up) to be the preferred method of changing channel. Figure 16 (d) shows participants preferred to pause the TV with the raise and hold gesture (forceful raise of palm in stop motion at TV, forceful sweep downwards to resume). Figure 16 (e) shows a clearly preferred interaction with the wave motion for setting a thermostat to away mode. Figure 16 (f) provides more of a mixed picture with twirl in the air and swipe up/down and twirl in the air and switch both ranked highly with the former the overall winner.

Results for the entire study can be found in the Newcastle University Research data store[63]. Each interaction ranking and calculated preference can be found here.

Follow-up Interviews

I found it interesting to reflect on some of the comments that were made during the study, with regards to the ways the participants felt themselves to be powerful or enabled by the system. Participants were generally very excited at the prospect of such a system to control devices in their home. *"It feels like magic"*, *"there's a performance to it"*, *"it seems quite fun"*. In introducing the motivation for this work, I posit that expressiveness and meaningful control was missing but sought in the smart home. The comments from our participants appear to validate this rationale. It is tempting to assume that our objectives are to elicit the most efficient and practical interactions with IoT devices, yet this disregards some perhaps more eccentric actions that users may prefer, as their performance engenders a sense of power.

Some participants had concerns about the complexity of some interactions, accidental control and accessibility – *"If it did all those things – I'd just get annoyed with it"*, *"it needs to be as simple as possible"*. Some participants felt the system was perhaps too invasive: *"I'm perfectly happy turning on a switch – it's functionality I don't need"*, though they found some of applications more compelling. The contextual awareness of such a system was an important consideration, *"when I'm watching the football and I clap I don't want the lights going off"*, *"confusing with actual human interactions could be a problem"*. Many of these comments reflect a clear observation with regards to expressive IoT

---

[63] Expressive Things Study Dataset, https://data.ncl.ac.uk/articles/Expressive_Things/12318683

interactions. Simple state changes and other simple device interactions are sometimes not appropriate. Increasing the complexity involved in controlling something is a common complaint about IoT device interaction and should be avoided in gestural IoT device control.

Some participants had concerns about some interactions being awkward or uncomfortable. *"Some of the wrist ones look uncomfortable"*, *"the interactions look tiring"*. Several participants also commented on the social acceptability of such interactions: *"I don't want to look like a tit"*, conversely though another participant commented, *"showing off to others would be cool"*. Participants were also largely positive when reflecting on the requirement of Expressive Things to be wearing a smart device on their wrist, *"some of the apps are so fiddly – this is more compelling"*.

Participants were largely positive about the system as a whole. Contextual awareness of such a system was a common theme amongst participants, both in terms of appropriate actions in response to a command and in avoidance of false positives. Adding expressiveness to an interface should also not necessitate the usage of such a system for control, control should still be possible through existing interfaces. Disconnection and deactivation, *"you need to have a clear understanding – when the interaction has stopped"*, accidental interaction after completion would leave users frustrated.

5.8. Rationale

We pose in the section "State of the Art" that the current state of IoT with manufacturer specific interfaces used to control individual devices is inadequate. As such, we look to use the elicitation study to help us identify an interaction set/language that transcends individual devices and interfaces. This is needed when considering the need for a wrist-worn sensor to control multiple IoT devices. In response, an interaction model was developed; to guide research and development of expressive IoT interactions. Based on these study findings, we have composed this model to capture what we consider to be the four key components of cross-device IoT control:

1. Thing Identification
2. Characteristic Selection
3. Control

4. Disconnection.

5.8.1. Thing Selection

Disambiguation of device control is a vital component of expressive interaction. Any feedback that follows the control of a particular device or set of devices must be clear. As such, identification of a single device or selection of a set of devices to interact with has been highlighted as the first stage of any interaction. The process of identification can be both implicit and explicit. Explicit selection could occur through making use of the other techniques (e.g. manipulating a physical interface or explicit device selection using voice). Implicit selection can be informed by the characteristic control, such as a wrist roll interaction when only the lights in the room are capable of responding to that gesture. I did not ask participants to consider device identification in our study, to reduce complexity in their interactions, with the exception of the select all devices interaction. This interaction included an explicit select all interaction for most participants.

5.8.2. Characteristic Selection

Devices that have multiple characteristics or attributes require a means to identify the desired one, or combine characteristics together for control. Again, this can be determined implicitly or explicitly. Explicit selection can occur again through a voice command or utilizing the characteristic control of a physical interface. Implicit selection of characteristics may occur when a characteristic is the only appropriate characteristic for an interaction, or the subset of controllable characteristics for a group of devices. Selection can also happen implicitly in response to a device-initiated change, the device that initiated a change can then easily be selected. Participants often used different interactions for different characteristics, with different degrees of freedom utilized. Some expressed a desire for a stepper to navigate through available characteristics.

5.8.3. Characteristic Control

How the control over a characteristic for a device or group of devices occurs. This can be considered as a dialogue, back and forth between device/s and characteristics, with feedback provided by the device characteristic change. The design of these interactions should consider the range and freedom of movement in this phase. In the study, participants preferred discrete interactions for simple state changes (e.g. wave for away mode on thermostat) and continuous interactions (e.g. wrist roll for lighting brightness) for value changes.

5.8.4. Disconnect

These interactions require a defined end, finalizing and disconnecting after applying control is further complicated by accidental continuation of control. As such, disconnect requires explicit consideration when designing interactions. An expressive interaction which fails to satisfy this requirement will struggle to provide the desired results for users. Explicit disconnection could occur through a voice command, or specific gesture, or implicitly after a stationary timeout. In the case of state changes, the characteristic is de-selected enabling further control after another characteristic selection. Participants, though not explicitly asked to do so, did consider how they would disconnect from a device. They overwhelmingly considered relaxation or dropping of the wrist down by their side as a preferred explicit disconnect.

5.9. State Diagram

To assist in the development of future gestural IoT interactions we have developed a state diagram. This diagram encapsulates the interactions produced through our interaction elicitation study and aims to address the issue of device and characteristic selection, which were not covered explicitly by participants in our study. I have adapted Buxton's tri-state model [14] of interaction to describe the different states possible during an expressive IoT interaction.

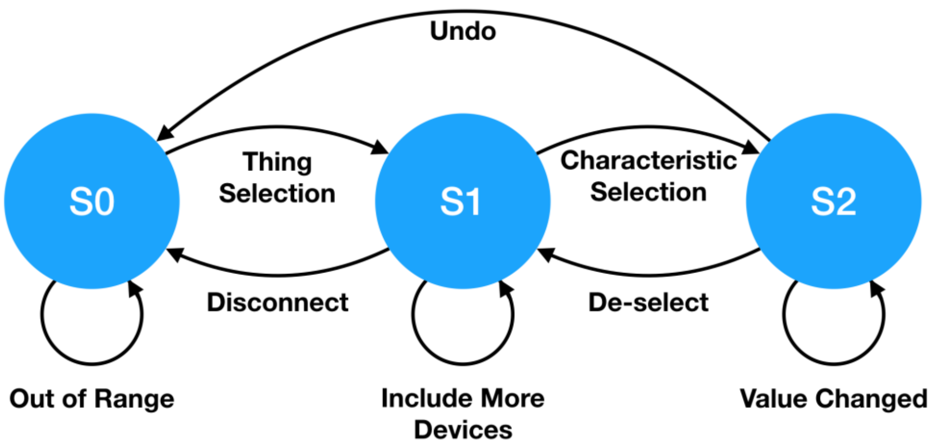


Figure 17: Expressive Things IoT interaction state diagram

To better demonstrate the suitability of this state diagram to capture expressive interactions, I will describe through a set of example interaction scenarios.

5.9.1. One Device, One Characteristic

For a device with one characteristic available for control, we start in S0, selection of the device moves to S1. As there is one characteristic available for control, characteristic selection can be implicit, in this case we implicitly move to S2. In S2 we are able to change the value of a characteristic, this can be a state change (e.g. on/off) or a discrete value (e.g. 0-100%). The state diagram adequately captures this scenario.

5.9.2. One Device, Multiple Characteristics, Less than DoF

Introducing more characteristics for control, however less than the degrees of freedom for control available. Starting in S0, selection of the device moves to S1. As we have less characteristics than available degrees of freedom we can again have implicit control, through initiation of interactions mapped one-to-one to the available characteristic (i.e., implicitly move to S2). Once in S2 we can control the selected characteristic value. We can move back to S1 through a deselect interaction, select a new characteristic for control and move back to S2. A user can then de-select a characteristic again, moving back to S1, then move back to S0 with a disconnect interaction. The state diagram adequately captures this scenario.

5.9.3. One Device, Multiple Characteristics, More than DoF

More characteristics than available degrees of freedom for control precludes implicit characteristic selection, as such we must select a characteristic explicitly, moving from S1 to S2. Starting at S0 a single device can be selected using a device selection interaction, moving the user to S1. Characteristic selection must be explicit, so a specific characteristic must be selected before being controlled in S2. The user can then de-select a characteristic, moving back to S1, select another or disconnect moving back to S0. The state diagram adequately captures this scenario.

5.9.4. Multiple Devices

Control of multiple devices of the same type or different types to control is also captured by this state diagram. Starting in S0, a user can select a single device to control moving to S1, then include more devices with more device selection interactions staying in S1. Selection of a characteristic to control follows the same principles as above, with number of available characteristics from the sub-set of device characteristics guiding whether selection can be implicit or explicit. Moving to S2 once all devices are selected, with a characteristic selection interaction the user can control a characteristic value on the group of all devices. De-selection of a characteristic moves back to S1, where another characteristic can be selected, or the user can disconnect all selected devices. The state diagram adequately captures this scenario.

5.9.5. Undo

Due to the nature of the expressive interactions being constructed there is a higher chance of misinterpretation. As such, I have included a method to undo the result of characteristic interaction. When in S2 a user can undo the most recent characteristic change that has been applied restoring the value to the original state. This undo interaction de-selects and disconnects the device/s returning the device/s to their original state before the last characteristic selection operation.

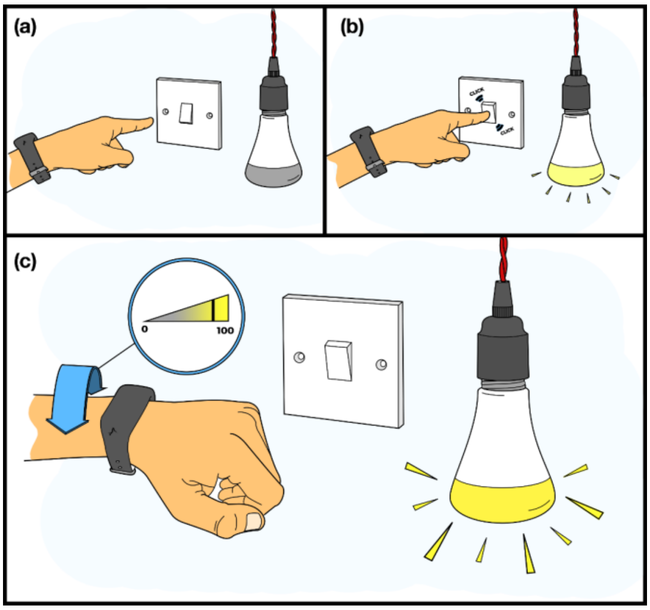


Figure 18: ET Lighting: (a) Light off, system in S0; (b) Switch on selects device moves system to S1; (c) Brightness control through roll of wrist moving to S2 implicitly; Disconnection occurs when wrist is dropped to side (relaxed).

5.10. Prototype Implementation: Expressive Things

In order to explore the range of interactions from our study and model that could be implemented using currently available hardware, I developed the Expressive Things prototype (ET). Building upon the interaction model, I produced a system to control lighting, music, media and temperature in the home. ET can be considered as a system that adds expressiveness to current IoT device interactions. Combining convenience with multi-device control in local communication contexts, ET utilizes existing hardware and interfaces to provide rich, expressive interaction for the home. Movements of the hand/wrist are tracked from a wrist-worn IMU with tracked changes being directly applied as control parameters to IoT device characteristics. ET can be used in conjunction with voice commands to provide context to a subsequent wrist-based interaction. ET enriches interactions with existing interfaces such as light switches or TV remotes providing access to typically 'out-of-reach' characteristics of IoT devices while looking to exploit interfaces with high levels of each immediacy, context and control. Combining interfaces to construct interactions with adequate coverage of each.

5.10.1. Technical Description

The Expressive Things prototype implementation uses an OpenMovement WAX9 IMU[64] streaming 3-axis Accelerometer and 3-axis Gyroscope at 50Hz, which is connected to a Raspberry Pi[65] over Bluetooth Low Energy (4.0). The Raspberry Pi mediates communication between the sensor and the IoT devices in the environment using Node.JS[66]. An estimation of the device orientation is calculated using Madgwick's sensor fusion algorithm AHRS [56]. This device orientation estimation is then resolved to pitch and roll angles as well as a calculation of linear acceleration. The system uses a simple gesture recognizer to track shaking, tapping and twirl in the air gestures. These interactions and data interpretations are directly applied to the control over relevant IoT devices. Focusing of the device to be controlled was achieved through an Amazon Echo Dot Voice Assistant[67]. Running a custom Alexa 'skill', users can focus a device to interact with by specifying a device followed by an appropriate attribute to control (*"TV Volume"*). LIFX[68] and Philips Hue[69] bulbs can be focused using the physical light switch associated with the device. Low-level network information is used here to provide timely state and network information about lighting, when a device disappears from the network and reappears, this is focused as the device for control. I consider events in an IoT space as any change of state, or detectable initiation of control. Toggling of a light

[64] OpenMovement WAX9, https://github.com/digitalinteraction/openmovement/wiki/WAX9
[65] Raspberry PI, https://raspberrypi.org
[66] Node.JS, https://nodejs.org
[67] Amazon Alexa Echo, https://www.amazon.com/echo
[68] LIFX Lighting, https://lifx.com
[69] Philips Hue, https://www2.meethue.com

switch, a voice command, activating a front door sensor, are all considered to be IoT events. In our interaction model terms, an IoT device event, either user or device initiated, moves the system to an S1 state for that device.

5.11. Expressive Things in Media

Expressive Things (ET) allows us to replicate the simplistic controls of existing media playback devices in a more convenient and comprehensive manner without modification of existing hardware. When considering Object-based Media playback ET can potentially provide the necessary expressivity to maintain control over more complex playback experiences. In the study for Expressive Things, I explore replication of the rudimentary media control systems offered by speakers and TV sets, providing play/pause, fast-forward and rewind. In doing so, I demonstrate a system that is arguably more immediate than the existing remote interface, as your wrist-worn sensor and voice are present-at-hand, but not very expressive – expression with such simplistic control is difficult and largely unnecessary. Object-based Broadcasting experiences envisage a more complex playback scenarios with a range of devices and playback mediums under control, play/pause, fast-forward and rewind no longer suffice.

Object-based Broadcasting workflows enable a new level of dynamism in media playback. The more nuanced configurations of the media potentially provide the opportunity for responding to more nuanced control mechanisms, enabled by a more expressive control system such as Expressive Things. Accompaniment – the system follows the users' lead – using explicit interactions to guide this complex Object-based experience and drive the narrative/sequencing. Building upon the infrastructural foundations offered by the Expressive Things project, I believe there is a compelling use case for more nuanced interactions to control Object-based experiences. Accompaniment, as it sounds, is envisioned as means to make content playback experiences more of a guided experience. Rather than a typical follow along media playback, the use is more involved, they can influence the content through decision arcs or other opportunities for reconfiguration.

Accompaniment style players – provide a very explicit and detailed control dialogue between the user and the player, could prove compelling Object-based Broadcasting experiences. In order to more clearly elucidate some of these accompaniment style experiences made possible by concepts such as Expressive Things I have produced some user scenarios.

5.11.1. Exaggerated Action

In many instructional or typical follow along media experiences, the consumer is dictated to at a particular pacing and expected competence. The presenter may caveat this with certain sections of the media dedicated to close-up walkthroughs of expectedly difficult sections. However, for the sake of brevity and broad appeal content makers will often tend to avoid a slow-paced show, instead opting for more content and encouraging the end user to pause as and when necessary. I have already discussed in detail the possibilities for reconfiguration to address some of these short comings and a more detailed feedback loop to the media playback, automatically detecting the stage the user is at ensures the media does not get too far ahead, adding detail if competency is known ahead of time. However, we can go beyond this when considering more nuanced control interfaces like Expressive Things. Take for example a section of a cookery show with a rather complex knife work task, filleting a fish. An experienced user could let this section play through, completing the task along with the presenter. However, if a user is struggling they could begin exaggerating the knife movement. This explicit accompaniment style interaction could inform the media playback engine that the current knife work task requires more detail. The content could then be reframed, or another shot selected. A close-up of the task with other supplementary data could be displayed, at a slower pace, with perhaps a more detailed voice over. This subtle but explicit interaction could provide a compelling accompaniment style experience to the end user. Providing a non-intrusive means to adapt the media to individual needs.

5.11.2. Dynamism in Input

Expressive Things introduces a new level of universality for control interfaces. Relying on its event-driven approach to input we can re-appropriate a particular control interface in an entirely new way. In the prototype discussions above, we have control of the brightness of a light with a TV remote volume, switching off all devices in a room (TV, stereo, all lights) by toggling any of the lights off in the room. I sought to restrict the usage of this technique so as to not introduce too much complexity for the end user. However, there are opportunities for dynamically creating input devices for content. In a more generic sense the way a user is currently interacting with objects in their home could influence the media. Consider the dynamism offered by some modern games, a user can select, and to some extent dictate, the narrative but also influence some more subtle characteristics of the game. Changing the lighting, moving objects in the scene. In each case the characters and narrative may be affected by these changes. Similarly, we could break out beyond the screen with accompaniment style interactions that influence the film in some way. This could be a subtle and passive reconfiguration, for example a cookery show swapping steps based upon detected ingredients or equipment. Or more fundamental shifts, changing the entire narrative by for example providing information to the character via a phone call. I have already discussed some of these possibilities previously, Expressive Things as a framework governing event-driven interaction provides the foundation for such work.

5.12. Summary

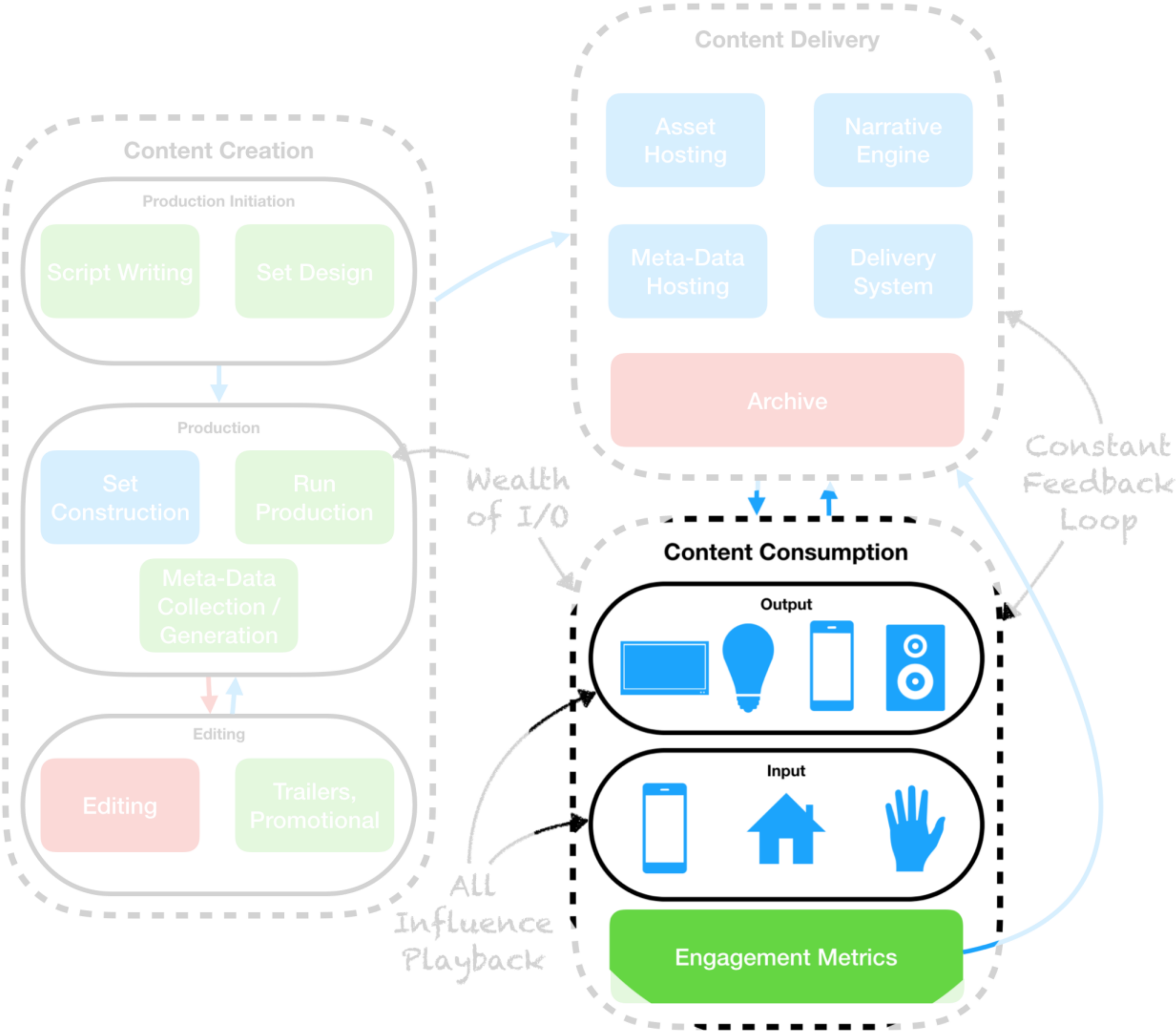
In this chapter, Maintainting Control, I present Expressive Things, a novel interaction system for smart home technology. Taking the current disparate and siloed nature of some Smart Home and non-Smart Home devices and exploiting the most favourable aspects of each to construct a “complete” control system. A control system that adequately covers the highlighted important properties of Immediacy, Context and Control. This project alone contained three significant user studies eliciting interactions, confirming the most favourable of each and testing the constructed prototype.

This chapter presents some of what I consider are fundamentals of the control systems required to provide a platform and ‘playground’ for Object-based Broadcasting experiences. Expressive Things demonstrates opens the door to bringing many typically siloed or otherwise ‘dumb’ control interfaces and objects into ‘play’ in the home. Instrumentation of individual objects with adequate labelling will remain infeasible for the forseable future. I believe strongly that a system such as Expressive Things provides the opportunity for some of Object-based Broadcastings more radical workflows, enabling wealth of information that can be feasibly captured in production to be adequately exploited in production.

I reason that such a control system is immediately applicable to Object-based Broadcasting. As the concept of Object-based Broadcasting is developed we have more parameters and degrees of freedom available to users. We need to develop control systems offer similar degrees of freedom. Expressive Things is a comprehensive example of this.

Sensed Player

Chapter Six



Open Lab at Newcastle University

BBC

Chapter 6. Sensed Player

In this chapter, I discuss the possibilities for sensor data driven playback experiences, exploiting similar ubiquitous computing technologies to those deployed in the Media of Things. In the chapter Media of Things, I explored the use of IoT technologies to generate time sequenced data during capture of video content in a studio environment. In the literature, I highlight the wealth of potential data/meta-data from a production that is lost, or inadequately exploited for Object-based Broadcasting. Similarly, post-production processes generate potentially useful data/meta-data, such as filtering out mistakes and tagging duplicate content.

Following on from the Media of Things deployment and exploring rich an expressive control with Expressive Things, I wanted to explore a novel consumption concept. The Media of Things dataset (Table 3) provides a rich source of descriptive data, labelled actions, object-tracking and raw signal data. Indeed, when considered in conjunction with some of the frameworks proposed in the chapter Maintaining Control, there is an opportunity to build a more nuanced and user-led consumption experience. I identified a potential avenue for the continued development of Expressive Things as a control system for Object-based Media. The contribution of Expressive Things as a mechanism for bringing more of the Smart Home into 'play' offers an intriguing proposition for a reactive playback system.

I sought to exploit the MoT dataset to develop a consumption experience which was novel and drastically different to existing consumption experiences, to test Object-based Broadcasting at its extremes. Playback influenced directly by, and indeed driven by, user interaction now seemed entirely plausible. As a first question, I asked: by deploying sensing technologies in the consumption environment, could we track enough about the user's interactions to inform the playback experience in some way?

Initially, I thought of the Sensed Player concept as a purely editorial experience. Going through the media as an initial pass for an editor or producer to familiarise themselves with the content – what better way to experience the media sets than to perform the actions again. This would perhaps enable editors to make a 'rough' edit of the media, mimicking the actions performed in sections of the media they want to look at a player could seek to the appropriate part of the film. This could provide a natural and fluid experience for editors, streamlining their initial workflows and immediately removing 'bad' sections of the film. Moving on from this as an editorial concept, this could also be envisaged as a playback experience. Focussing on the Media of Things dataset of a cookery show, more competent and familiar chefs would perhaps prefer less of a follow along experience, rather let the media follow along with them. Using an Object-based Media set such as that produced by the Media of Things production we could build an experience that followed the user along, cutting into the relevant parts of the media based upon detected interactions.

6.1. Crowd Influenced Playback

Broadcasters use detailed user analytics to provide invaluable information about their audiences. These are used when comminsioning content, targeting certain demographics and in content delivery. Object-based Broadcasting provides an interesting prospect for playback experiences themselves to be influenced. If we can liberate ourselves from the constrains of pre-packaged media, we can reconfigure the media at the player as we choose. This reconfiguration could take the form of crowd, or rather consumer influenced playback; taking again the example of a cookery show from MoT, as consumers decide to cook the elements of the recipe in a different order can this influence the original media?

User attentiveness is a well-researched area of HCI in media, investigating what works well in terms of audience immersion with eye tracking [21] and other forms of engagement tracking [39]. In novel consumption mediums such as VR, the BBC has been looking at generating heatmaps of user interaction [10], tracking where a user is looking throughout a film. This project was conducted in order to better understand user's interaction with VR films, however it is also being applied to aid content delivery. Streaming high-resolution 360-degree film to a user's headset is wasteful if they only view certain portions of the film. Approaches to adaptive streaming have looked at predicting user movement, enabling low resolution streams of certain portions which are less likely to be viewed and reducing the data rate of the film [51].

A user led playback experience, also nicely lends itself to consumer generated edits. As this was initially an editorial concept, there is potentially a nice symmetry in design here. Developing a player which is similar to its editor enables consumers to develop edits for a show simply by playing along. This could provide an interesting crowd-sourced editing process for content. The initial experience is determined by a production team then, over time, this is influenced by the community and refined into a more popular take on the show.

6.2. Immersion

Immersion in media is a fundamental quality of consumption, the level of immersion is often directly related to the user's enjoyment and the dissemination of information. Immersion is also a largely subjective feeling, both in terms of the user and the subject of the content. In a drama show, immersion could be considered as the level to which the user feels present in the show or taken out of their world and into another. This is emphasised by how we consider Virtual Reality experiences to be almost the pinnacle of immersion – the user feels as close as possible to being enveloped in the narrative of the show.

However, when considering more instructional films like a cookery show this description of immersion could be found wanting. Rather than a desire to be transported into the experience, it is perhaps reasonable to want the experience to be fluid. Traditionally, such experiences would need to have been: pause; rewound; or sections skipped, hardly a favourable experience for the end user as they inevitably try to keep up with Jamie Oliver's "20-minute meals" …

In Object-based instructional or follow along experiences we have a new opportunity to realise this immersion – this is what I wished to explore with the Sensed Player. Can I construct a fluid and immersive experience that suits media with defined segments or follow along type experiences? Setting out to construct an experience that is rather radical, where the media follows the user, if we can realise this extreme of matching intention with media, can we support the states in between as well? In doing so construct a compelling and immersive Object-based media experience.

6.3. Building upon Expressive Things

In the chapter Maintaining Control, I introduce the concept of Expressive Things. A framework for more nuanced and detailed control over devices through combination of existing interfaces to IoT devices. This system provided exploited favourable characteristics of voice, gestural and physical interfaces to develop a higher bandwidth communication medium for IoT. The qualities of Immediacy, Context and Control were discussed as fundamental to the Smart Home, but lacking in existing interfaces – in Object-based Media experiences such qualities are also desirable. Immediacy in control over media playback, beyond play/pause/fast-forward/rewind moving to specific sections of the media. Context with media adapting to suit its environment, input from sensors in the environment providing appropriate control. Finally control, exploring new ways to manipulate the media playback, driving the experience with deep intuitive and powerful control – engendering a sense of mastery over the media.

In building upon the foundations of Expressive Things, Sensed Player explores a radically different method of playback of media. Focussed upon follow-along media experiences, I investigate allowing the user to direct the playback through performance. The player lets the user take the lead, exploring an implementation of accompaniment style interactions, the playback system accompanies (follows) the user through the experience.

6.4. Sensed Player

In order to explore this concept of a user led consumption experience, I needed to ensure that I had a test bed that enabled the most dynamic/extreme reconfiguration of the media. Reflecting on the dataset from the Media of Things production we had levels of data from abstracted inferences to detailed raw sensor data.

To build a Sensed Player, I needed to achieve a similar understanding of the consumption state as we had in production. I decided to work from the highest level of inferences (e.g. chopping board in use in area x). Inferences are matches from sensory hardware in the consumption environment. This enabled a more abstracted implementation; the production sensory infrastructure does not need to match the consumption infrastructure. Such a system could remain sensing agnostic, while also allowing for graceful degradation of the experience as sensors are removed.

The Sensed Player flips the traditions of follow-along media consumption on its head – it follows along with the user, dynamically switching to the detected segment of the media. This results in a deconstructed narrative where the experience is guided entirely by the user. To move to a different segment of the media they simply start making a different part of the recipe.

6.4.1. Sensor Data

In Media of Things, we offer a wealth of sensor data for generation of data at the point of capture, this data provided a range of inferences and contextual information to the media. This sensor data provides detailed descriptive data describing the underlying media assets. To build a sensed playback experience we need to exploit this sensor data to construct a means to accurately infer where a user is within the show.

6.4.2. Technical Description

The Sensed Player is a web-based player supported by a range of sensors, mediated by a message queue server. Raw data from the environment is supplied to a central hub for subsequent processing. This emulates an equivalent coordinated Smart Home environment where sensors would be coordinated through a gateway service such as Apple HomeKit[70] or Google Home[71]. The web player streams data for use in matching the playback to current performed actions in the form of raw sensor data and recognised activities, depending upon the media sequence. The playback head is moved based upon the Sensed Player's inferred position.

6.4.3. Architecture

The Sensed Player was built around the Media of Things dataset, as such the sensors have been chosen to be appropriate for detecting actions for this dataset. However, I will describe the more generic application scenario.

Object-usage Information

OpenMovement WAX3[72] accelerometers were placed in all utensils, pots and pans – these sensors provide basic movement information. Objects related to the consumption experiences beyond a cookery show could similarly be supported through placing movement sensors in these objects; coordinated with the objects taxonomy of actions, this could be used to infer action information. Accelerometer streams from the WAX3s

---

[70] Apple HomeKit, https://apple.com/ios/homekit
[71] Google Home, https://home.google.com
[72] OpenMovement WAX3, https://github.com/digitalinteraction/openmovement/wiki/WAX3

are broadcast via a proprietary version of Zigbee to a receiver node. The receiver nodes, running Windows 10 IoT[73], unpack the data and forward it to the RabbitMQ server.

Recognisers read the raw sensor data from the message queue and run noise filters on the signal to generate usage information. These inferences are handed back to the message queue for use by the Sensed player. Detected action information can be used to match against similar actions being conducted in the playing media.

Object-location

RFID tags on objects were used on items within the space. These tags provide contextual information to the Sensed Player. In the cookery scenario ingredients placed or moved at a particular time can be used to identify the current step in the recipe. RFID pads continuously poll for changes, the status of the pads is fed into the RabbitMQ server. Similarly, recognisers are run on the dataset cleaning up the noisy transponder signals.

6.4.4. Sensed-Playback Matcher

This data along with the appliance data straight from an energy signature monitor (Smappee[74]) is used to match against sections of the media. Time-sequenced meta-data from the film is directly matched at the action level (inferred from accelerometery) and objects in use (inferred from RFID).

---

[73] Windows 10 IoT, https://developer.microsoft.com/en-us/windows/iot

[74] Smappee Energy Signature Monitor, https://www.smappee.com

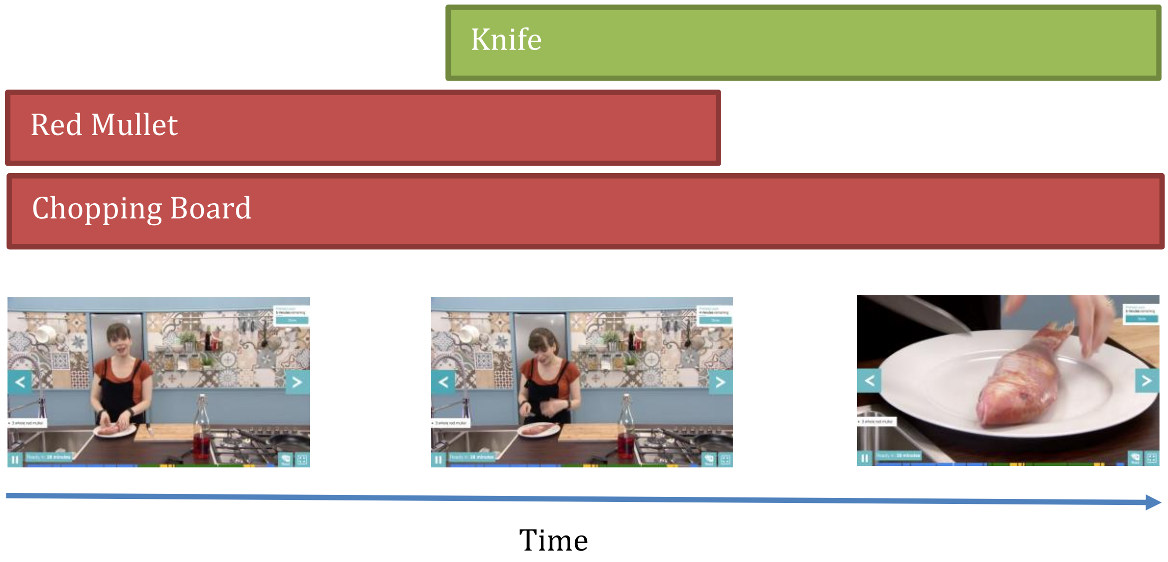


Figure 19: Representation of MoT dataset.

In Figure 19, we can see the level of detail available from the MoT dataset even without inferences. In green we have detected knife movement from the knife's accelerometer in the handle. In red we have the items detected on the bench from the RFID pads. Utilizing this information, we can construct a map of possible entry points into the media. In playback with the possibility of tracking similarly detailed levels of interaction data, we can navigate the media by moving the playback head to a matched location.

6.5. Segmentation

In order to match sequences of the show against real-time data we may need some level of show segmentation. Bringing the user in at the right times, at the beginning of a detected piece or at the right subsection could be crucial to maintaining immersion in the media and avoiding potentially jarring mid-section breaks. In the Sensed Player I chose to dynamically segment the media. Sections of the film are matched against detected actions; these actions are amalgamated into sections based upon configurable timeouts. A configurable lead time to a segment is also available.

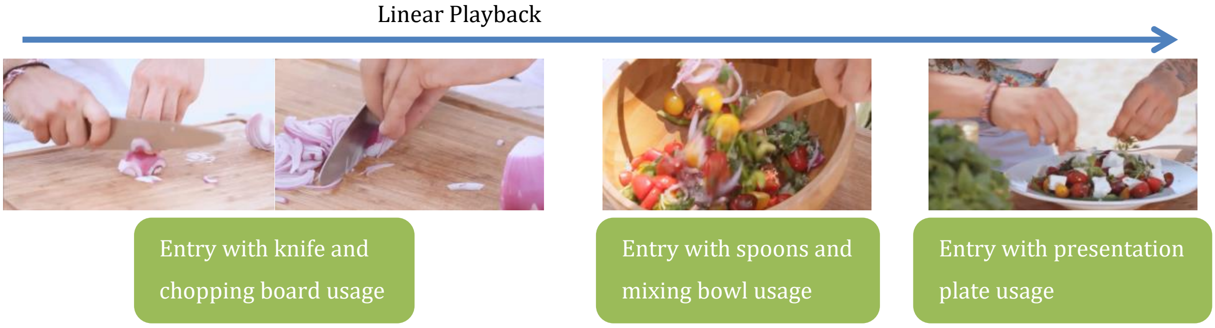


Figure 20: Segmentation of media, entry points and relevant matched actions to enter.

There is a trade-off here between the depth of reconfiguration, and potential for breaking immersion through placing the user in the middle of a sequence. For this reason, the dynamic segmentation offered by the Sensed Player can be considered a first pass. Providing an initial set of breaks in the media, a simple visual editor of the segments could be provided to allow an editor to tailor these segments to suit the film.

6.6. Visualisation

In an experience such as the Sensed Player, a user can enjoy a fluid and user led playback experience. Content (depending on the output medium) can be presented based upon the matched segment of the media as described above. This type of playback experience does present a new set of challenges to the user. How do they know what is to come next? How should the narrative be guided without user input? How should the user know what options (in more branching narrative experiences) are available? How does the user know how to get to a certain section of the media through performance?

These questions can be accumulated into what is natural to the user and what may be considered un-natural. In a cookery show, with simplistic output like the recipe text to follow; a user could naturally assume what is necessary to start the next step perhaps. In more complex or unbounded experiences this may seem less clear. This is where clear visualisations are required.

If we consider other novel consumption mediums like Virtual Reality research and industry has moved towards more explicit prompts: marking sections of media as interactive elements; displaying more detail; or moving the media forwards in some way. In branching narrative experiences, a user can select a pre-defined set of paths through the experience. In these user led experiences, which offer more nuanced and potentially complex OBB paths through media, we need to be careful. Explicitness arguably leads to a broader appeal, more users are able to navigate the media and have a more informative experience. The experience is therefore more guided, and similar in many regards to a branching narrative experience. However, this will undoubtedly be at the expense of immersion. Jarring experiences where the next step is explicitly shown to the user, or the options are presented could break any immersion for the user as I outline above. Therefore, there is a trade-off between explicitness and immersion.

6.6.1. Developing a Solution

In developing a solution to this problem of visualisation of the media, I looked at existing approaches to visualisations of higher levels of film. Digital video editors have long used the tried and trusted timelines with previews of the media shown alongside the associated audio and text track. iMovie[75] and Adobe Premiere Pro[76] have pioneered these interfaces as a means to quickly navigate long media tracks to seek to the desired point. Online media players have begun experimenting with crowd sourced engagement metrics [115] to show users the most interesting portions of the media. Indeed, Periscope briefly experimented with automatically generated highlights of livestreams based on engagement/interaction metrics [117]. Taking inspiration from detailed labels on media players, can we exploit the detailed labelling of the dataset from Media of Things to generate a more detailed contextually labelled player?

6.6.2. Chronology out the window

Traditional media consumption experiences across almost all domains, rely on one constant, chronology of the film. In OBB experiences, this ever-present constant is gone. Instead, we have a reconfigurable media source whose playback sequencing and length can be influenced by a wide range of factors. This puts most of the players, and the traditional media presentation tools which focus on this centre point of time out of the window, I wanted to explore something new. Jackson *et al.* produced an interesting visualisation of an entire film with a system called Panopticon [45,67]. This player enabled a viewer to see an entire film at once with the film cut up into segments that are

[75] Apple iMovie, https://apple.com/imovie
[76] Adobe Premiere Pro, https://adobe.com/premierepro

tiled and presented as a scrolling view across the screen. This novel concept inspired me to build upon this idea with the Sensed Player.

Instead of a timeline of media, I decided to work on an interface that shows a representation of the current detected state. When in an unknown state the player shows each segment of the show, alongside a visualisation of the entry points to this media.

Figure 21: Sensed Player in operation.

6.6.3. Accompaniment

When constructing this prototype, I wished to focus on an experience that was radically different from existing experiences. The purpose of this being to test Object-based Broadcasting in some of its more traditional and radical scenarios, exploiting the opportunities presented by the rich Media of Things dataset I generated. The Sensed Player flips traditional follow-along media experiences on their head, quite literally. The player follows the user, working to detect where a user is within a media set and presenting the media through the most appropriate medium. This user led experience, is however, a bi-directional communication when considering the visualisations presented to the user.

As such, we can consider this experience as a dialogue in playback between the user and the Sensed Player. The user is leading the experience, but the options and feedback from the visualisations and media being presented to them can influence their interactions. These types of interactive experiences are analogous to research around attentiveness with branching narrative [17,68,91,93]. To further develop this bandwidth for communication between the two parties, the Sensed Player introduces some more explicit interactions to guide the experience – I call these interactions accompaniment interactions. They provide more explicit guidance to the player, providing a means to select the desired clip for playback.

6.7. User Study

In order to evaluate the performance of the Sensed Player I ran a user study. The purpose of this study was to investigate the desirability of a reconfigurable playback experience such as the Sensed Player. In order to assess the player, I recruited 10 participants (avg. age 28.2) to use the Sensed Player in our kitchen studio with similar instrumentation to the Media of Things shoot.

6.7.1. Methodology

To provide a more compelling experience, more akin to the developed and closer to market Sensed Player system, I edited together sections of the media and made markers for the relevant portions of the media. Rather than use our still in development end-to-end Sensed Player experience and matching raw signal, I pursued this more abstract approach to provide a less error prone experience to users. We wished to obtain qualitative data about the experience, not about the systems performance as it currently stands. I also ran a stripped-down version of the Sensed Player itself without the overall video visualisation and instead provided paper recipe steps to guide the user.

Each participant was provided a set of ingredients and two short recipes to follow. There were roughly 10 steps in each, the preparation steps could be done in any order, the recipes in any order. The participants had the system explained to them, and were encouraged to experiment with the interface (moving objects and see how the media responds). The first recipe was for a Greek salad, this recipe had several vegetables for chopping and various mixing and presentation steps – all steps in the recipe were on film, and the content was fast paced. The second recipe was for tiramisu, this recipe had several pre-prepared sections, chosen such that the media would typically get ahead of the user. This film was much slower in pacing and the ordering of steps slightly more important.

The constructed Sensed Player implementation seeks to a specific point in the media when certain ingredients or utensils were moved. These points in the media are pre-defined as the beginning of segments, the participant can select a new section for playback by beginning the next stage of the recipe. While participants were working with a live Sensed Player, detecting their actions and responding accordingly, there was a significant manual process driving the experience. Manually overriding mistaken actions and moving the media to the relevant section when not detected. I wished to evaluate the experience not the Sensed Player's current implementation. When participants took longer than the media playback section to complete a task the task would loop until they had finished. The recipes took on average 45 minutes to complete. At the conclusion of the study, I conducted a semi-structured interview with participants, covering their experience of the player and discussing the appropriateness of such a player in their home in a variety of contexts.

6.7.2. Results

Each of the participants were overwhelmingly positive about the Sensed Player as a consumption experience – "this is how cookery shows should always work". I was encouraged by comments such as, "bit of a unknown discovery element to it, which is kind of fun … It was just really fun". This demonstrates a user interest in such reconfigurable experiences. Several participants commented on the lack of clarity over the next steps, with no way to explicitly tell the player to continue, besides just starting the next task and letting it follow you along this was sometimes confusing to users – "I would like to have sort of known the next step".

Customisation

A question I asked each participant was to consider the appropriateness of such a player, with varying degrees of personal competency, or familiarity with the task. For example, a recipe they have cooked for years with the Sensed Player acting more as a suggestive system, providing alternative means or flavours to a dish. This provoked an interesting discussion with one participant around the appropriateness of such a suggestion, she was concerned with how that would fit into her carefully controlled diet – "I think that would just annoy me. I know I can make that better by adding olive oil or something, but I can't". We discussed at length the possibilities for such a system for someone with complex dietary requirements, or individuals just trying to be healthier. The Sensed Player in its current form presents a pre-selected media file, or set of media files, to the user as they cook along with it. The current experience is explicit, selecting this recipe to be completed, it doesn't adapt to the environment but simply follows the user along. However, a more implicit and passive experience could also be conceptualised. Rather than working from a particular recipe, instead we could work from a higher level; suggesting alternative methods of preparation of ingredients from a range of content for example. Providing a more assistive experience but also providing the potential for spontaneity media consumption, with more contextually aware content consumption systems.

Music & Dialogue

All participants discussed the use of music or pacing of dialogue in their reflections. Three participants talked about the music specifically. The Greek salad video had high tempo and rather annoying music. The participants noted that this contributed a significant amount of stress to this recipe, feeling as though they had to keep up with the pace of the show – "The music really stressed me out", "I hated the first one, the music was really annoying". Furthermore, the participants noted that the choice of music in the Greek salad film made it incredibly obvious when a switch occurred in the media, leading to a more jarring experience.

When developing media for a player such as the Sensed Player, content creators should be careful in their selection of music, and the parts of a show to use it in. Due to the depth of reconfigurability offered by the Sensed Player (enter at any point, switch at any point) this could be difficult to eliminate entirely, but a single piece of high pace music throughout should perhaps be avoided.

6.7.3. Further Personal Reflections

There are several things of note to discuss from my personal reflections of running through the Sensed Player with our participants. Looking at the reconfigurable aspect of cookery shows as a concept, there is a fundamental unfamiliarity here that was prevalent throughout the study. People by nature coming into the study were expecting a recipe and instructions, naturally they appeared to feel like there was an order to the instructions. Whilst they were encouraged at the beginning of the study to explore making the recipes in their own order, and experiment with the system, most did not. Overcoming the natural instinct to follow instructions, in the order they are presented, appeared to be difficult. The Sensed Player was intended to differ from a traditional stage detection experience, as the entry and exit from the media was not governed by a dependency map of steps to complete. The users could enter and exit from the any point in the media simply by beginning a certain step. However, the player in its simplistic form presented here, does fall into the trap of being a stage detecting system when used with a set of ordered steps.

Furthermore, the expectation that people will use the intended tools, equipment and ingredients is a complete falsehood. For this study, I used the same instrumented kitchen environment as we had for the Media of Things shoot: utensils with wireless accelerometers, bowls, ingredients and equipment with RFID tags, and appliance usage through an energy signature monitor. Now, in a professional production a chef would have already decided upon their equipment for each task. Using an unintended item can be trivially corrected in post-production. No such niceties exist in a live playback, much like its live media production, we are at the raw unfiltered edge of sensor data. Regularly participants would pick up one item when they meant to choose another, use a different, un-instrumented spoon or fork when more "appropriate" tools were available. Therefore, relying upon some semblance of consistency across users, or even within the same user is a rather hopeless endeavour. This was to some extent unexpected, I had assumed there would be at least some symmetry between the chef and user. This is perhaps a factor of a follow along experience that must be considered, if the user initiates the playback then they cannot see what they "should" be using. A workable Sensed Player must go beyond a selection of objects being used at once to match segments, or the user will end up confused. However, keeping such a system could provoke an interesting dialogue between the media and the user encouraging them to make their tasks easier by using appropriate tools which are available to them.

6.8. Summary

In embarking upon this prototype's development, I had a stated goal of exploring an extreme form of reconfigurability of media using raw sensor data. The aim of which was to understand to what level we could offer a completely non-liner media experience. At the same time such an experience, ignoring some of the typical constraints of for example continuity, had the opportunity to provide some more compelling Object-based experiences, without significant modifications to the production workflow. A production team could perhaps simply perform their piece in an instrumented studio, with sufficient instrumentation to provide machine "readable" context. Subsequent playback within similarly instrumented environments could apply simple inference mapping to control the playback head.

In conducting the study above, I adapted two existing pieces of content by adding stage markers and detectable actions to start them. This provided mixed results overall, stages were not as clearly defined as I would have liked, segments were sped up or removed for brevity and the shot changes were abrupt and jarring. The first video of the Greek salad, was highlighted as problematic by several participants, it was very fast paced and used a clearly expert chef with rather irritating music. The second video of Tiramisu was better; however, some steps were almost blended together, making some of the stage detected switches jarring. This was somewhat intentional with the selection of the media, I wished to explore both adaptation of existing media sets, and how comprehensive the media content needed to be. My results in this endeavour were encouraging however, a little tainted. Some parts of the recipes had too little opportunities for true reconfigurability. Generally, participants did not appreciate the high pacing of the Greek salad video nor the music, they enjoyed the other film but were sometimes visibly confused by the lack of detection driven movement by the player.

In this study, I purposefully chose to ignore some broadcasting fundamentals. Namely continuity and linear narrative. The rationale behind this was to enable the deepest form of reconfigurability, if there are no defined entry or exit points, there can be no continuity, no linearality to the experience whatsoever. However, as with much of my research thus far, there is a clear balance/set of trade-offs to be considered. As the depth of reconfigurability increases, so too do restrictions on the content need to be put in place. Continuity was not too much of an issue in the Greek salad recipe, however in the

Tiramisu recipie performing steps out of order resulted in bowls and utensils being used that were not necessary and the user could get confused. Similarly, with the music, the lack of a linear arc to the experience meant that music made the experience jarring. Balancing the “quality” of the content against the potential for reconfiguration will be the important consideration for experiences like the Sensed Player. For example, a film which is meant purely as an assistive piece, or suggests different ways of doing things could have no music and little concern for continuity. The result being that if such a piece was watched in a linear player it may seem quite basic and dull.

Discussion

Chapter Seven

Chapter 7. Discussion

In setting out the scope of this thesis, I focussed my research on Object-based Broadcasting and the many barriers to the commercialisation of this concept. I reasoned that these barriers required a fundamental rethink of the production workflow, advocating for a more modern and data integrated process. Any solution or steps towards a solution to the challenge presented to the industry would require work at both ends of the content production pipeline. Content makers had already set out their stalls with grand visions of what they believed Object-based Broadcasting could be. I looked to investigate how to make those grand visions a reality, exploiting the explosion in Smart Home technology and the Internet of Things more generally.

In this chapter, I embark upon detailed discussions at the culmination of this work. Each of these subsequent discussion topics are based upon my personal experiences in executing Object-based productions, engineering experiences, and running case and user studies of these technologies.

7.1. Specificity vs Genericism

In outlining the motivation behind this research, I posited that current explorations of Object-based Broadcasting and more generally responsive/reconfigurable media experiences have focussed on a pre-defined set of output mediums and consumption scenarios. In order to realise the true universality of the concept of Object-based Broadcasting, we needed to develop technical solutions to enable the deployment of generic media assets appropriate for an unknown set of output mediums that could befall the media. Looking at what has gone before and what I view as possible, I have discovered it to be predominately down to a careful balance of specificity of the technical solutions supporting a broadcast and the potential genericism of its output possibilities. Indeed, I will now argue that such a balance, or rather a trade-off, exists in almost all parts of the production pipeline.

Media of Things as an implementation maintained its genericism by being sensor agnostic; it could accept any text real-time data stream and store this information for playback. Simplistic categorisation of this data based on its contribution to, for instance, the area of interest classifier, meant that with little configuration, sensing units could be added. However, anything beyond simplistic inferences – such as “this responding means this is in use” – requires specific recognisers to be built, which could be applied in a range of contexts but inherently increases the specificity of the underlying system. The level of obtainable detailed data/meta-data from sensing within the production studio is therefore directly proportional to the engineering time to construct this data. This carefully struck balance in developing data/meta-data to build compelling experiences, contrasted with the pre-production engineering, is one that which content makers must consider.

Expressive Things provided a framework to build control interfaces to smart home devices, combining existing interfaces, and exploiting the most favourable aspects of each to construct detailed and nuanced controls. Constructing consumption experiences for a piece of content may also entail constructing control scenarios and recognisers. Experiences may be more compelling but could require more specific engineering of elements of the consumption workflow. For example, the Sensed Player required a high degree of explicit engineering to construct the experience. The depth of reconfigurability available did not need constructing but the sensor mappings and some of the high-level meanings of certain events did (this RFID transponder means chopping board). Some of these engineered components have replicability, for example in other cookery shows. However, this is the intrinsic balance that I discuss here; engineered solutions will inevitably need re-engineering for improvements or as technology progresses. The suitability for constructing more reconfigurable and responsive experiences will inevitably come down to budgetary constraints.

7.2. Redefining Creativity

Working at the bleeding edge of any transformative concept will inevitably lead to challenges for existing roles. Artificial Intelligence, many believe, will lead to a requirement for Universal Income [97] as skilled and unskilled jobs are superseded by machinery. Even in this work, I have proposed automating aspects of the production workflow entirely, such as camera operator automation and assisted editing. However, there is another fundamental challenge to the industry with Object-based Broadcasting: when the experience of a user is no longer uniform, how can content makers impress their creative flair upon their content? In a show such as HBO’s Game of Thrones, the directors and producers have carefully controlled each aspect of the consumers path through the narrative. These carefully-crafted audience experiences are a cornerstone to these tremendously successful TV shows and films. Object-based Media potentially introduces a vast array of unknowns for these content makers to consider. What will the ordering be? Which output mediums are going to be used? How will the user respond to the media? Each of these questions raises fundamental affronts to their otherwise end-to-end controlled experiences. The media is no longer pre-packaged and is instead presented in response to a whole host of environmental factors.

Considering this disruptive evolution in content production, I view these prophases not as challenges but as opportunities. Object-based Broadcasting experiences provide a new level of immersion and the opportunities for creativity, rather than being lost, is shifted to the logic model backing these experiences, how the media responds to the environment, and how the environment responds to the media.

The fundamental driver of the creativity, I believe, is in 'intelligent' response to environmental stimuli for the media. I refer to this as the logic model, the mechanisms through which a certain action or output medium is responded to. This redefining – or more appropriately reframing – of the creative aspects of content production opens up exciting possibilities for content makers. An analogous situation presented to an industry would be open-world games, whereby narrative paths are becoming less structured and more responsive. A gamer has more freedom to explore a game as they please. Logic models backing such games are in many ways similar to ones backing OBM, though their input and outputs are less well defined. In developing these OBM experiences in such a way, I believe we can provide a new definition for what we consider to be creativity in content production.

In discussion of the many aspects of media production, we contemplate the automation of certain sections. Much of the data labelling I generated with the Media of Things sensor suite were things that were non-trivial to label, such as: movement of objects identification of which objects were in the scene at any one time, or the movement of the chef's wrists. However, when we discuss moving beyond these measures into more complex automation of tasks that relate directly to creativity – such as camera work and editorial decisions – I view these proposals as purely assistive systems; curation is still very much the domain of man, for now. In summation, media remains a fundamentally creative process; we are simply taking the depth of creativity to new levels.

7.3. Taming the Wild West

Being presented with a concept such as Object-based Broadcasting was as many parts exciting as it was daunting. At the outset, much of the conceptual work backing it seems fanciful and farfetched – relying on technical advances in AI, sensor fusion and in-home data integration that seemed simply unachievable. However, I soon came to realise that the power of the concept of OBB is in its flexibility, both in terms of its ubiquity of applications but also its graceful degradation. As such, I decided to approach my research into supporting the creation and consumption of Object-based Media through three very different case studies in different areas of the production workflow and to differing degrees of realising the generic concept of OBB. In doing so I obtained invaluable insight into each part of the workflow and the contribution of IoT across the array of challenges facing the broader adoption of OBB.

Object-based Broadcasting relies on revolutions at several points in the media life-cycle; any suggestion of a technology or group of technologies that could aid in its development needs to be suitably tested – assessing the suitability at the extremes of reconfigurability, production/consumer sensing ubiquity and infrastructural sensing. Media of Things first and foremost was developed as a means to an end. We deployed an excessively instrumented kitchen workspace in a production studio. This was not necessarily the most efficient implementation of Media of Things but was developed in order to maximise the amount of contextual information capture, as well as determining which sensors are most appropriate and useful. Such a deployment enabled me to work on a wide range of potential OBM experiences as well as investigating some of the more nuanced aspects of automated/assisted editing. In this sense, the MoT dataset provided the most generic base for building a whole host of playback experiences which Object-based covers.

Moving on from the Media of Things deployment, I investigated the other extreme of Object-based Broadcasting with the Sensed Player. Using a similarly instrumented kitchen environment, I looked to explore a media experience with no defined entry or exit points – the ultimate form of reconfigurable media where the environment influences every aspect of playback. Through a user study, I wished to understand: a user's reaction; what changes, if any, were required to be made to the underlying media; and how an environment detecting and following you along could work as a consumption experience.

Deploying technologies like CryptoCam, I sought to explore the potential gold mine of information already collected about ourselves and our environments – developing technology to encourage a more open approach to data sharing while maintaining some semblance of privacy.

Finally, with Expressive Things I sought to reign in the control over these novel content experiences with a system to integrate smart home control events with wrist-worn gestural control. In doing so, I exploited the most favourable aspects of each current interface to the Smart Home.

Each of these case studies and developed prototype technologies were purposefully very different. At the outset, I outlined my methodology in relation to this research to investigate the feasibility of Object-based Broadcasting at each stage of the media lifecycle. As such, high level deployments of these technologies were conducted at each of these stages, thereby providing a comprehensive analysis in terms of end-to-end work but perhaps not in terms of breadth of production scenarios.

The "wild west" I refer to here relates to the potentially unbounded set of possibilities for reconfiguration of media with Object-based experiences. In this research, I aimed to bound (tame) this conceptual work through these prototype deployments.

7.4. Immersion vs Accessibility

Throughout this thesis, I have discussed the possibilities for increasing immersion with Object-based Media. Immersion in an activity – be that a cookery show or a tense thriller – is a quality that content makers strive to achieve. In a constrained environment, such as a cinema, immersion is almost guaranteed; every part of the experience is crafted by the film makers and emphasised by the cinema owners. Immersion with Object-based Media can also be attained through compelling experiences that perhaps follow you along, such as the Sensed Player. Media is your orchestra, you are the conductor. Interactive experiences, where the media moves beyond the typical single screen experience and towards a more environmental immersion, also illustrate the potential of OBM in this regard – each object that is moved, or each action of the consumer could be used to influence the playback in some way. While many would agree such experiences are desirable, they present new barriers to inclusion and accessibility. Losing control of the playback experience was indeed such an important problem that I wished to dedicate an entire chapter to addressing some of the issues it presented; how can we ensure that we design such experiences to limit potential accessibility barriers that may arise? This is where I believe there is another clear trade-off – immersion and accessibility. Rather, I should say, immersion for those without accessibility issues. Increasing the complexity of the experience will lead to more accessibility issues; moving around the room to continue a narrative, or even making an Expressive Things interaction to control the experience in some way, will introduce barriers for some.

Accessibility is often an afterthought when considering innovation in the technology sector. Media inherently is designed for the sighted and non-deaf individual; adaptations of existing media to be more accessible to those who are not typically take the form of post-production audio description and subtitles. However, there are varying degrees of disability and accessibility needs which are simply ignored with current broadcast outputs, instead lumping all together as one. As an example, someone with photosensitive epilepsy is simply told not to watch certain shows or broadcasts, despite the fact that these scenes could be easily adjusted to remove these potential hazards. A person with learning difficulties may find some dramas inaccessible; complex plot lines and subtle references to previous episodes or films may leave them with a half-baked experience. Coming back to audio description and subtitling, users may want varying degrees of both. For example, a partially sighted person may want slightly less detailed audio description. A partially deaf person may only want hushed voices to be subtitled. In a traditional media setup, many of these problems are difficult to address, with little data to provide machine readable context. Object-based Media provides a new opportunity for deeper integration of accessibility services and a wider variety of accessible outputs. Thus far, I have discussed reconfiguration and responsiveness of OBM with regards to creating more compelling experiences for those without accessibility needs. However, we can support more tailored experiences for these groups too. For example, a drama which: reconfigures to add flashback segments to the show's subtler cues; bright flash sequences could be removed; or a cookery could present adjusted steps for those with relevant disabilities.

Accessibility and media have historically shared a fundamentally antagonistic evolution. Object-based Media provides the opportunity to breach new levels of content immersion and expressive control; however this must be carefully considered as a trade-off with accessibility. Furthermore, conceptually, Object-based Media was meant to lead to more tailored and responsive experiences. This tailoring could and should also take the form of more accessible media for all.

7.5. Everything as a Sensor

In the chapter Maintaining Control, I discuss a conceptual control system called Expressive Things. In discussing the theory behind the concept, I discuss the possibilities for using IoT events as input to the control system. In the chapter Integration, I present the CryptoCam concept and discuss the opportunities for cameras as a powerful sensing unit, not only providing context to other datasets but with ever-advancing Computer Vision techniques providing detailed sensory information. Across these propositions is an underlying belief in 'Everything as a Sensor' – exploiting typically underutilised datasets which, in isolation, provide little useful information but that, through aggressive integration, can support a wealth of complex experiences – from smart home device control with Expressive Things, to complex reflective films with CryptoCam. Many systems record a wealth of information about their usage, and these datasets are often largely meaningless to the end user. However, through integration of these datasets, we can use them to drive other experiences – Event-driven Programming if you will, although at a higher level. The concept of Expressive Things is an example of a fundamentally event driven experience – exploiting IoT event data to create control experiences.

-aaS is a popular suffix when referring to cloud and hosted services. It refers to the process of migration of systems into more generic standardised hosted services, potentially allowing for greater integration of these systems. In discussing the possibilities for sensing in supporting Object-based Broadcasting, we have largely ignored the infrastructure to support this, instead assuming such a system's presence and offloading all sensor processing and handling to remote nodes which the majority of existing systems have proffered. Centralisation of data and processing, however forwards the siloeing of data; which I argue against in this thesis. Integration of services at the cloud end only offers a limited and high latency solution in the commercial interests of larger organisations. In order to realise some of the experiences I discuss in this thesis we need a fundamental refocussing of this approach, namely local processing and secure sharing of data perhaps based on contextual or locality-based security measures. This was the reasoning behind the CryptoX concept – the sharing of data between services that may wish to utilise it. These kinds of distributed processing and sharing of data of "potential interest" I believe is crucial to providing some of the more complex Object-based Broadcasting experiences in homes. Manufacturers are working towards similar integration of services at a more fundamental level. The Thread Group[77], of which Apple and Google are members, discuss network level solutions to integration of services, creating a comprehensive solution for the Internet of Things in the home and beyond. Advances such as these could provide the required hardware integration to enable low latency deep media interaction with devices in the home.

7.6. Machine Understanding

Throughout this thesis I have referred to the term "Machine Understanding"; widely used in the Machine Learning & Artificial Intelligence community, it refers to a state beyond Machine Readability, where a computational system responds with an input or set of inputs with an appropriate response. Crucially, this response is typically not governed by a simplistic set of arbitrary logical steps but rather a more nuanced "understanding" of the series of inputs and perhaps its influence of the response. Obviously, without a true and complete AI, we cannot apply typical humanistic characteristics to our work. Take for example the human brain; we can (largely) get all the understanding we need from just our eyes of, for example, a cookery show. Computer Vision techniques to achieve similar context are, as I have already outlined, insufficient. They lack the detailed nuances of certain actions and the detailed description of objects. This is where with Media of Things we sought to introduce new inputs to any such logic models or intelligence. Through providing explicit sensing, we construct a logic model that is Artificial Intelligence in its most basic sense. The wealth

[77] The Thread Group, https://threadgroup.org

of data available to us at this level meant that we needn't go any deeper – simple, easy-to-define IFTTT models sufficed. This also meant that some elements of the engineering of complex narrative experiences could be propagated up to editors without much retraining.

Machine Understanding does not necessarily need to be complete in the case of media production labelling or narrative engines; a minimum viable product type approach should be considered here, for example, with the Media of Things dataset captured from a cookery show, the understanding required to develop some compelling OBB experiences can be quite limited. Indeed, the Sensed Player worked on a simple matching of detected actions against other detected actions in playback. In order to automate sequencing of shots, or identify the most salient portion of a film, one could build logic models from the data. For example, if a utensil is moving or equipment is in use then most likely a detail shot is required, and the region of focus is the location of the appliance or equipment. This does however come back to the balance of Specificity and Genericism; building a model such as this is highly specific to the dataset and the type of production. Advances in Artificial Intelligence could provide the answer to ingesting data from sensors in productions and environments and producing useful output. Quite simply, more input sources produce a greater Machine Understanding and therefore more useful output.

Conclusion

Chapter Eight

Open Lab at Newcastle University

BBC

Chapter 8. Conclusion

In this chapter, I discuss some of the possibilities for future development of my work, building upon the wealth of research on Object-based Broadcasting experiences and Smart Home control. I will also reflect upon my research question and objectives, identifying answers to my research question and highlighting the contributions to this answer from each objective.

8.1. Future Work

Object-based Broadcasting is a broad and far reaching concept that has the potential to revolutionise content creation and consumption. In this thesis, I investigated a means to realise some of the opportunities this concept affords for broadcasters and consumers. I did this through building solutions that: support the construction of media assets required for reconfigurable and responsive media; and support the consumption of reconfigurable media in the home exploiting existing smart home technology. My research in this area, whilst answering the questions I set out to answer, has opened up a raft of potentials avenues for further exploration. In this section I will outline some of these possibilities for future research.

8.1.1. Media of Things in new Contexts

The Media of Things collaboration we ran in conjunction with the BBC focussed on a cookery show, whilst the capture system remained sensor agnostic, the sensors selected were targeted to common cooking tasks. Also, inferences engineered from the dataset were often highly specific to cookery shows. A kitchen is a constrained environment that typically contains smart devices in the modern home (oven, voice assistant etc.). Media of Things should be explored in other constrained environments, for example studios offer some of the same niceties of a kitchen: a fixed space; indoor; typically specified seating; and designated zones for certain tasks. Such an environment could be highly conducive to the existing Media of Things capture suite and inferences. Crucial to the future development of Media of Things is the development of a generic model to sit on-top of the captured data, abstracting sensors output to a system of generics that broadens their immediate reapplicability. For example, accelerometers are usually used to detect movement, the raw signal data, in most use cases, provides little extra to the set of constructed inferences. Abstracting such a sensor to simply a movement sensor could enable editors to place sensors on objects when this is all the information they need, such as a car in a scene or an item for sale on a telemarketing show.

Another avenue for exploration is integrating existing captured metrics into OBB narrative engines and logic models. For example, Formula 1's position indicator may identify salient portions of the films i.e. overtakes. Similarly, in a tennis match the serve power or length of rally could inform automatic edits, with highlights being automatically constructed around these rich descriptive data sources. Increasing the depth of sensing in sporting events coverage could also provide compelling metrics to the consumer too. Take for instance a boxing match, placing a sensor in the gloves could enable fatigue tracking, punch power measurements and tracking of statistics like missed punches. The ever-increasing reliability, reducing size and price of sensing hardware opens up lots of exciting opportunities for research when coordinated through systems like Media of Things.

8.1.2. Taking Auto-Directors Beyond Vision

Automatic or assisted shot editing, framing and sequencing are a continuing focus of media researchers. Ingesting raw media assets and developing broadcast ready output is the dream. However, as I reason from the outset of this research, we need to move beyond what can be reconstructed with Computer Vision techniques. Intricate detail, that often cannot be detected through video analysis alone, can be detected with explicit sensing as demonstrated in the Media of Things and Sensed Player chapters. Each new sensor input gives us new features to work from, developing more complex and accurate models to work with. Indeed, the Area of Interest classifier employed a simplistic logic model to identify salient portions of the set, with camera framing performed with pre-configured regions from each camera. Through integrating sensor and vision data we could potentially provide the missing ingredient to realise the auto-direction dream. Far from removing the creativity from the film making process we have an opportunity to create assisted editing tools, democratising film creation further and streamlining professional resources to make more content.

8.1.3. Sensed-Editor

Building upon the Sensed Player chapter, such an interface could work for editors. Exploring their captured media through repeating steps in the show, navigating the media in a natural and dynamic way. The Sensed Player followed the consumer along as they performed a set of tasks, this provided a compelling experience for our participants. They commented that it felt like a more intuitive and natural way of using how-to or typical follow-along media experiences. If we consider the raw output from a media production, or rather a data rich production like the one in Media of Things, an editor is presented with a wealth of information with indexing is typically based on simple scene markers, retake markers and timecodes. Navigation and crucially familiarisation with this data can be a time-consuming process. A simple adaptation of the Sensed Player could enable an assisted editor, or media familiarisation system. To view the footage for a particular part of the show an individual simply needs to mimic the tasks conducted at that point in the media. When used in conjunction with some of the accompaniment style interactions discussed previously, one could envisage such an interface providing a first-pass edit. Marking usable parts of the media, selecting which cameras are appropriate for each scene etc.

8.1.4. Smart-Editor

Capturing a wealth of contextual information from each media asset opens up new possibilities for the creation of tools and services to support the editorial process. Extensions to existing media editing tools to support rich time-sequenced data could provide compelling productivity enhancements to existing editors. For example, if an identified shot contains an imperfection, or mistake; the section could be selected in order to be replaced. At this point the editor could search the captured raw media for similar clips, matching based on detailed data/meta-data. For instance, a scene showing the chef preparing some vegetables on the chopping board, a slight pause by the chef results in the clip is too long. Highlighting that section of media to be replaced presents all other clips where the chopping board, vegetables and knife were in use. The editor can select the desired clip and replace it easily.

8.1.5. Actuations of the Smart Home

I have discussed and constructed experiences in this thesis that exploit Smart Home technologies for controlling and informing the playback experience. There is however a large and ever-growing network of home devices that could be utilised for playback: moving beyond 5.1 or 7.1 audio to room-based playback; using lighting to create ambient effects; or turning on an air-con unit to create a physically chilling effect. These could be considered actuations of the home, making physical changes to the playback environment to accompany the playback experience. In doing so, I believe we can construct a more compelling consumption experience under the umbrella of Object-based Broadcasting.

8.1.6. Expressive Things 2.0

Expressive Things opened up a vast array of possibilities and opportunities for control over IoT devices in the home. Through construction of a cross-device interface I demonstrated how such an approach can prove fruitful to achieving control that is immediate, contextually aware and rich. An interesting development to Expressive Things would be to include more ambient, less specific information. I discussed in the chapter Maintaining Control, the possibilities for responding to planned events, for example dismissing a planned heating switch off; initiation of control was not necessarily user-based. Similarly, ambient sensors like Peripheral Infrared (PIR) sensors are often used to initiate actions, such as setting off an alarm or just turning on a light in a room. Expressive Things could be expanded to include such a sensor, using it to initiate control, users could then express their true intent to manipulate a characteristic of a room (e.g. control brightness of lights upon entry with writst roll), or perhaps dismiss an automated action all together.

8.2. Reflecting on Objectives

- Objective One: Run an Internet of Things sensor instrumented production, to establish suitability of the Internet of Things in media asset labelling
- Objective Two: Explore greater integration of existing production, personal and infrastructural data to provide greater understanding to generated film
- Objective Three: Exploit Smart Home and Internet of Things devices in the home to crate immersive content experiences

In culmination of my work I should reflect upon my objectives I presented in the introduction. I have already highlighted answers to these questions throughout this thesis in the relevant chapters. I will use this section to clearly address each objective in full, highlighting contributions of each objective to the research question and develop some discussion around future research directions.

Objective One: Run an Internet of Things sensor instrumented production, to establish suitability of the Internet of Things in media asset labelling

The chapter Media of Things realised this objective. I ran a production in collaboration with BBC Research & Development with a deeply instrumented studio. At the outset I would like to state, that my analysis of the suitability of IoT in productions was conducted in specific contexts – namely cookery shows. Media of Things proved successful in labelling media assets for this type of production. Conclusions beyond constrained productions, such as cookery shows or other studio-based shows, require further research. In the MoT production, I was able to adequately label interactions and performative action within the production in near real-time provided there was clear configuration of sensors ahead of time. This is crucial, configuration as a pre-production (can be post for non-live) is, to some extent, the limiting factor – this comes back to the point of specificity vs genericism.

We can construct compelling Object-based Media experiences with IoT sensing as demonstrated by Media of Things and for example the Sensed Player. However, more research is needed into different production types. The diverse nature of media production means that similar sensing infrastructure needs to be tested in other studio type productions and in more open, less constrained productions. The system architecture I developed, being sensing agnostic, should enable Media of Things to remain as the foundation to these productions, but exploring what sensing is appropriate and what contextual information needs to be captured is key.

Objective Two: Explore greater integration of existing production, personal and infrastructural data to provide greater understanding to generated film

Integration of existing datasets was a significant focus of this work. Rather than collecting more and more data I investigated how existing datasets could be exploited to create new inferences. In the chapter Integration, I presented two technologies designed to encourage more open data in sports and open data sharing practices in CCTV. Hawk-Eye EDGE-SENSE used an Open-Source sensor (and dataset) to aid umpires in a cricket match; data was maintained in its raw format for later use and video footage contextualising this bat data was also stored. CryptoCam demonstrated a practical means of large-scale sharing of video footage, a technology such as this provides new opportunities for integration of life logging datasets; this could also further develop the idea of a camera as a sensor, providing greater understanding and context to these typically siloed datasets. The subsequent discussion provoked by CryptoCam provided a vital insight into some of the issues and consequences of greater integration. Video footage of an office or public space may reveal little of interest, integration with other datasets has the potential to provide invasive detail of an individual's life. There is a balance to be struck between potential benefits and privacy implications.

Furthermore, I believe that enabling the small-scale production of Object-based Media relies on making more from what we have. Practically speaking, small production studios and independents are unable to replicate a production like our BBC R&D shoot. However, I believe there is potential in integrating consumer hardware and other personal datasets, we can create some compelling labelling of captured media to enable some level of reconfiguration. As it currently stands, we are going to struggle to create something like for instance the Sensed Player, but we could highlight the salient portions of a media set using information such as high heart rate, for example.

Objective Three: Exploit Smart Home and Internet of Things devices in the home to crate immersive content experiences

The modern home is a fertile landscape of input and output mediums. Object-based Broadcasting provides an opportunity for media to spread beyond the bounds of a TV screen. In the chapter Sensed Player, I experimented with a radically different playback experience – exploiting the sensing hardware present in an instrumented environment to track a user and let the media follow them. I demonstrated that such an experience can be compelling and intuitive to users, however content needs to be selected/filmed appropriately to suit the experience. Furthermore, controlling existing media experiences is simplistic, fast-forward, rewind, play/pause can be provided by any number of command interfaces, physical remotes, voice etc. This suitability is lost when considering some Object-based experiences. For example, with Sensed Player, its playback head is influenced by a wide range of factors (detected action, stages left, etc.), we could classify these as degrees of freedom afforded to the consumer. As such with Expressive Things I looked at a broader concern, as we increase the characteristics under control with smart home devices more generally how can we create control interfaces which keep up and offer the required degrees of freedom to provide adequate control. In constructing Expressive Things I addressed the lack of IoT control interfaces that were immediate, contextually aware and offered complex control. Consequently, I also addressed the potential lack of control over the new degrees of freedom presented by OBM experiences.

8.2.1. Research Question

How can the Internet of Things and Smart Home Technology support the production and consumption of Object-based Media?

Media of Things answered the question of the suitability and contribution of the Internet of Things to the production of Object-based Media – sensing provides a practical and effective means of labelling actions and object usage in studio productions. Similarly, Smart Home technology and platforms to coordinate control over IoT hardware in the home provide a bedrock for building expressive control interfaces (Expressive Things), or coordinating sensor driven playback (Sensed Player). More open approaches to data sharing and integration of IoT sensor data also provides more detailed contextual information to captured film, or conversely captured film provides contextual information to data (CryptoCam).

The contribution of IoT and Smart Home technologies in supporting Object-based Broadcasting, is therefore significant. However, at the culmination of my work I would go further, I suggest that one is intrinsic to the other. In productions detailed data can be generated from IoT sensing in production studios and on sets, Media of Things [98] along with previous research such as Red-Tag [7] and BBC Primer/IPStudio [11,113] can be used to construct detailed labellings of contextual information that realise a wealth of complex Object-based Broadcasting experiences. I should balance this against the fact that the majority of my work was from a sensing perspective, as this is my background. However, it is undeniable that more information, from sensing or otherwise, enable more experiences.

In creating immersive experiences for users, Smart Home technologies provide a new level of potential. Predicated on appropriate data from the scene and logic models developed by production teams we can exploit lighting, appliances, unrelated device interfaces and more – now, one could argue that this is just a coincidental symbiotic relationship. The natural synergy generated from advancements in both smart home technology will undoubtedly assist in broadcasters drive for more immersive and reconfigurable content experiences. They are suitable partners for now, but may not always be as larger technology companies exercise their control and wall off their gardens.

8.3. Contributions of this Thesis

In conclusion of this thesis I would like to highlight and emphasise the contributions, as I view them of this work. This work is situated in the space of media, though nuanced and expanded with IoT and Smart Home technologies, constraining the work across two very disparate and seemingly disconnected mediums was a challenging endeavour. I approached this research agenda by attempting to tackle the distinct areas of media works, content creation (both professional and consumer), content delivery, consumption experiences and control. I feel I have delivered compelling research in each of these areas.

8.3.1. Content Creation and Consumption

Media of Things broached the question of professional productions with rich data labelling. Deploying feasible sensors and an architecture that scales I produced a system that reliably and accurately records previously inaccessible contextual information from professional productions. In the Sensed Player works, I further emphasised the importance of such a dataset in constructing a novel consumption experience. The depth of reconfigurability offered by such a playback experience would not be possible without such a data capture system available in production. As discussed in these chapters, manually labelling each potential entry and exit point would be infeasible. The chapter Sensed Player details a means for content delivery, both descriptive data and video that is extensible and has the potential to scale. The constructed prototype along with the user study provides compelling insight into some of the more extreme applications of the concept of Object-based Broadcasting.

8.3.2. Integration and Exploitation of Existing Datasets

In the chapter Integration, I explored the opportunities for deeper integration of existing personal and infrastructural datasets to create greater inferences about recorded media and other data. I explored a novel means of sharing data from fixed place cameras with CryptoCam – exploiting the typically untapped dataset of CCTV footage using existing personal data sources as meta-data to index the film. I contribute a patentable concept for the sharing of large amounts of data securely to those nearby. Delving into the societal impacts of CCTV, I also generated a conversation within our research group around the appropriateness of cameras in workspaces, raising the question of whether classical literature's reasoning around the "unobservable observer" no longer being the most pressing resistance to public cameras. This, while a by-product of my work rather than the focus, provided some guidance for future research in public cameras and more open approaches to data sharing in communities.

8.3.3. Novel Control Mechanisms for IoT and Media

Expanding on some of my previous work (Expressy [100]) I developed Expressive Things. This built upon Expressy's hand orientation estimation around a touch point on a tablet or smartphone. Instead decomposing IoT device interaction into events, combined with gestural control and exploiting favourable characteristics of existing interfaces. Through exploring this system for control in Smart Homes I also sympathetically demonstrated a compelling control system for media. Increasing the degrees of freedom available for control, I discuss how this can be naturally mapped to the increasing complexity of Object-based Media. As more devices, objects and screens are brought into the fold by this novel consumption medium we similarly should look to increase the degrees of freedom for control to match. Expressive Things provides complex and intuitive control without obstructive instrumentation of the person or environment.

8.3.4. Open Source Works

Throughout my PhD research I have produced a wealth of source code. Where appropriate (and legally able to do so) I have packaged and Open Sourced these works. I have worked on a range of novel technologies and systems, which I have often had to build out tools and services as they were either non-existent or poorly documented. I have made each of these tools Open Source and linked in the appendices below.

Each of the solutions I have engineered in my research, from prototypes to more complete solutions, I built in an extensible manner. In cases where it was appropriate/feasible, I went a step further to document the code and usage.

I strongly believe in the Open-Source movement; I have endeavoured to publish all components in a usable manner. I hope others will be able to exploit my systems, building upon them. I will be working on many of them in future as part of my continued research.

As an aside to the contributions behind this thesis, I have also been deeply involved with the continued development of the OpenMovement sensing project here in Open Lab, Newcastle University. I have contributed directly to the development of an Open Source sensing project the AxLE. I have greatly enjoyed developing tools and hardware throughout my research project, as and when we needed something better we built it. A thoroughly refreshing approach to research.

8.4. Final Remarks

In this thesis, I present a comprehensive analysis of the integration of Internet of Things and Smart Home technology into media productions and the media lifecycle more generally. I worked for the past three years at each stage of media workflow, building technological solutions to the feasibility issues surrounding the realisation of Object-based Broadcasting. I have contributed novel technical solutions and a range of study-based findings relating to IoT and media production.

I can conclusively say that IoT technology is a powerful tool for media productions. Sensing in studios themselves provide a wealth of previously inaccessible information, with subtle and ever more reasonably priced sensing. The data produced by such systems provides the required foundations to construct some novel consumption experiences, like those outlined in this thesis. What was previously limited by the reach of Computer Vision we can now augment with explicit sensing, providing reliable context capture to media productions. In consumption, as the complexity of visual media experiences ever increases so too will the complexity and pervasiveness of the Smart Home. There is a natural synergy between making our homes more intelligent and our constant yearning for more immersive and compelling media experiences. Exploiting IoT devices' new characteristics, voice interfaces dynamism, and other objects in the home – alongside ever more rich data and meta-data from productions – we can realise the exciting possibilities of Object-based Broadcasting.

References

1. Accenture. 2014. State of the Internet of Things Study from Accenture Interactive Predicts 69 Percent of Consumers Will Own an In-Home IoT Device by 2019. Retrieved August 10, 2017 from https://newsroom.accenture.com/industries/systems-integration-technology/2014-state-of-the-internet-of-things-study-from-accenture-interactive-predicts-69-percent-of-consumers-will-own-an-in-home-iot-device-by-2019.htm

2. Amr Alanwar, Moustafa Alzantot, Bo-Jhang Ho, Paul Martin, and Mani Srivastava. 2017. SeleCon: Scalable IoT Device Selection and Control Using Hand Gestures. *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*: 47–58. https://doi.org/10.1145/3054977.3054981

3. J. Assfalg, M. Bertini, C. Colombo, A.D. Bimbo, and W. Nunziati. Automatic extraction and annotation of soccer video highlights. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, II-527–30. https://doi.org/10.1109/ICIP.2003.1246733

4. Till Ballendat, Nicolai Marquardt, and Saul Greenberg. 2010. Proxemic interaction. In *ACM International Conference on Interactive Tabletops and Surfaces - ITS '10*, 121. https://doi.org/10.1145/1936652.1936676

5. R.M. Baños, C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, and B. Rey. 2004. Immersion and Emotion: Their Impact on the Sense of Presence. *CyberPsychology & Behavior* 7, 6: 734–741. https://doi.org/10.1089/cpb.2004.7.734

6. Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei, and Michael I. Jordan. 2003. Matching Words and Pictures. *Journal of Machine Learning Research* 3, Feb: 1107–1135.

7. Tom Bartindale, Daniel Jackson, Karim Ladha, Sebastian Mellor, Patrick Olivier, and Peter Wright. 2014. RedTag. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video - TVX '14*, 19–22. https://doi.org/10.1145/2602299.2602303

8. Tom Bartindale, Guy Schofield, and Peter Wright. 2016. Scaffolding Community Documentary Film Making using Commissioning Templates. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2705–2716. https://doi.org/10.1145/2858036.2858102

9. Tom Bartindale, Guy Schofield, Peter C Wright, and Peter Wright. 2016. TryFilm: Situated Support for Interactive Media Productions. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 1410–1420. https://doi.org/10.1145/2818048.2819929

10. BBC. 2017. BBC Taster launches virtual reality app. *BBC*. Retrieved from https://www.bbc.co.uk/mediacentre/latestnews/2017/bbc-taster-virtual-reality-app

11. BBC R&D. Nearly Live Production. Retrieved from http://www.bbc.co.uk/rd/projects/nearly-live-production

12. Richard A. Bolt. 1980. "Put-that-there." *ACM SIGGRAPH Computer Graphics* 14, 3: 262–270. https://doi.org/10.1145/965105.807503

13. Alex Butler, Shahram Izadi, and Steve Hodges. 2008. SideSight. In *Proceedings of the 21st annual ACM symposium on User interface software and technology - UIST '08*, 201. https://doi.org/10.1145/1449715.1449746

14. William A. S. Buxton. 1990. A Three-State Model of Graphical Input. *INTERACT '90 Proceedings of the IFIP TC13 Third Interational Conference on Human-Computer Interaction*: 449–456. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.139.1700

15. François Caron, Yee Whye Teh, and Thomas Brendan Murphy. 2014. Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes. *Annals of Applied Statistics* 8, 2: 1145–1181. https://doi.org/10.1214/14-AOAS717

16. Pablo Cesar, Dick C A Bulterman, and A. J. Jansen. 2008. Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content. In *Lecture Notes in Computer Science (including*

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 168–177. https://doi.org/10.1007/978-3-540-69478-6_22

17. Pablo Cesar and Konstantinos Chorianopoulos. 2007. The Evolution of TV Systems, Content, and Users Toward Interactivity. *Foundations and Trends® in Human-Computer Interaction* 2, 4: 373–95. https://doi.org/10.1561/1100000008

18. A. Churnside, M.E.F. Melchior, M. Armstrong, M. Shotton, and M. Brooks. 2014. Object-based broadcasting - curation, responsiveness and user experience. In *International Broadcasting Convention (IBC) 2014 Conference*, 12.2-12.2. https://doi.org/10.1049/ib.2014.0038

19. Tony Churnside. Object-Based Broadcasting - BBC R&D. Retrieved from http://www.bbc.co.uk/rd/blog/2013/05/object-based-approach-to-broadcasting

20. Henry Cooke. The Inspection Chamber. Retrieved August 29, 2018 from https://www.bbc.co.uk/rd/blog/2017-09-voice-ui-inspection-chamber-audio-drama

21. Anna L. Cox, Paul Cairns, Nadia Berthouze, and Charlene Jennett. 2006. The Use of Eyetracking for Measuring Immersion. *Proceedings of CogSci'06 Workshop*.

22. Jasmine Cox, Rhianne Jones, Chris Northwood, Jonathan Tutcher, and Ben Robinson. 2017. Object-Based Production. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '17 Adjunct*, 79–80. https://doi.org/10.1145/3084289.3089912

23. M. Davis. 2003. Editing out video editing. *IEEE Multimedia* 10, 2: 54–64. https://doi.org/10.1109/MMUL.2003.1195161

24. Deloitte University Press. The Internet of Things in the oil and gas industry. Retrieved from http://dupress.deloitte.com/dup-us-en/focus/internet-of-things/iot-in-oil-and-gas-industry.html

25. Boon-lock Yeo, Minerva Yeung, Bede Liu Dennis Yow. Analysis And Presentation Of Soccer Highlights From Digital Video. Retrieved June 2, 2016 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.24

26. Hannah Devlin. 2017. UK needs to act urgently to secure NHS data for British public, report warns. *The Guardian*.

27. P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth. 2002. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. . Springer Berlin Heidelberg, 97–112. https://doi.org/10.1007/3-540-47979-1_7

28. Andrew Garbett. 2017. Designing Community Driven Participatory Platforms. Newcastle University.

29. Gartner. 2016. Gartner Says Worldwide Wearable Devices Sales to Grow 18.4 Percent in 2016. Retrieved August 10, 2017 from https://www.gartner.com/en/newsroom/press-releases/2016-02-02-gartner-says-worldwide-wearable-devices-sales-grow-18-percent-in-2016

30. Benjamin J. Goold. 2002. Privacy rights and public spaces: CCTV and the problem of the "unobservable observer." *Criminal Justice Ethics* 21, 1: 21–27. https://doi.org/10.1080/0731129X.2002.9992113

31. S. Gregory. 2010. Cameras Everywhere: Ubiquitous Video Documentation of Human Rights, New Forms of Video Advocacy, and Considerations of Safety, Security, Dignity and Consent. *Journal of Human Rights Practice* 2, 2: 191–207. https://doi.org/10.1093/jhuman/huq002

32. John Guiver and Edward Snelson. 2009. Bayesian inference for Plackett-Luce ranking models. In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, 1–8. https://doi.org/10.1145/1553374.1553423

33. Edward T. Hall, Ray L. Birdwhistell, Bernhard Bock, Paul Bohannan, A. Richard Diebold, Marshall Durbin, Munro S. Edmonson, J. L. Fischer, Dell Hymes, Solon T. Kimball, Weston La Barre, J. E. McClellan, Donald S. Marshall, G. B. Milner, Harvey B. Sarles, George L Trager, Andrew P. Vayda, and Andrew P. Vayda. 1968. Proxemics [and Comments and Replies]. *Current Anthropology* 9, 2/3: 83–108. https://doi.org/10.1086/200975

34. Pam Hanley and Rachel Viney. 2001. Pressing the red button: Consumers and

digital television. *Cultural Trends* 11, 43–44: 35–60. https://doi.org/10.1080/09548960109365165

35. Richard Harper. 2003. *Inside the smart home*. Springer.

36. Martin Heidegger. 1977. The Age of the World Picture. In *Science and the Quest for Reality*. Palgrave Macmillan UK, London, 70–88. https://doi.org/10.1007/978-1-349-25249-7_3

37. Alex Hern. 2017. Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act

38. Alex Hern. 2017. Google DeepMind 1.6m patient record deal "inappropriate." *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/may/16/google-deepmind-16m-patient-record-deal-inappropriate-data-guardian-royal-free

39. Javier Hernandez, Zicheng Liu, Geoff Hulten, Dave Debarr, Kyle Krum, and Zhengyou Zhang. 2013. Measuring the engagement level of TV viewers. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2013*. https://doi.org/10.1109/FG.2013.6553742

40. Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex. Bulter, Gavin Smyth, Narinder Kapur, and Ken Wood. 2006. SenseCam: a retrospective memory aid. *UbiComp 2006: Ubiquitous Computing* 4206: 177–193. https://doi.org/10.1007/11853565

41. Jesse Hoey, Thomas Plötz, Dan Jackson, Andrew Monk, Cuong Pham, and Patrick Olivier. 2011. Rapid specification and automated generation of prompting systems to assist people with dementia. *Pervasive and Mobile Computing* 7, 3: 299–318. https://doi.org/10.1016/j.pmcj.2010.11.007

42. Clare J. Hooper, Patrick Olivier, Anne Preston, Madeline Balaam, Paul Seedhouse, Daniel Jackson, Cuong Pham, Cassim Ladha, Karim Ladha, and Thomas Plötz. 2012. The french kitchen. In *Proceedings of the 2012 ACM Conference on*

Ubiquitous Computing - UbiComp '12, 193. https://doi.org/10.1145/2370216.2370246

43. Chuanping Hu, Zheng Xu, Yunhuai Liu, and Lin Mei. 2015. Video structural description technology for the new generation video surveillance systems. *Frontiers of Computer Science* 9, 6: 980–989. https://doi.org/10.1007/s11704-015-3482-x

44. Simon Hughes. 2016. New app could ensure club players have nowhere to hide. *The Times*. Retrieved from http://www.thetimes.co.uk/article/simon-hughes-new-technology-could-ensure-club-batsmen-have-nowhere-to-hide-8s9l0kflh

45. Dan Jackson, James Nicholson, Gerrit Stoeckigt, Rebecca Wrobel, Anja Thieme, and Patrick Olivier. 2013. Panopticon. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, 123–130. https://doi.org/10.1145/2501988.2502038

46. Robert J K Jacob. 1995. Eye Tracking in Advanced Interface Design. *Virtual Environments and Advanced Interface Design*: 258–290.

47. Lorenz Henric Jentz, Nicolas Denhez, Young Duk Song, and YeongKyu Yoo. 2016. Augmented reality input device. Retrieved September 7, 2017 from https://www.google.com/patents/USD795256

48. Rene Kaiser, Marcus Thaler, Andreas Kriechbaum, Hannes Fassold, Werner Bailer, and Jakub Rosner. 2011. Real-time Person Tracking in High-resolution Panoramic Video for Automated Broadcast Production. In *2011 Conference for Visual Media Production*, 21–29. https://doi.org/10.1109/CVMP.2011.9

49. Rene Kaiser, Marian F. Ursu, Manolis Falelakis, and Andras Horti. 2015. Enabling Distributed Theatre Performances through Multi-Camera Telepresence. In *Proceedings of the 3rd International Workshop on Immersive Media Experiences - ImmersiveME '15*, 21–26. https://doi.org/10.1145/2814347.2814351

50. Rene Kaiser, Wolfgang Weiss, Manolis Falelakis, Spiros Michalakopoulos, and Marian F. Ursu. 2012. A Rule-Based Virtual Director Enhancing Group Communication. In *2012 IEEE International Conference on Multimedia and Expo*

Workshops, 187–192. https://doi.org/10.1109/ICMEW.2012.39

51. Kashyap Kammachi-Sreedhar, Alireza Aminlou, Miska M. Hannuksela, and Moncef Gabbouj. 2017. Viewport-adaptive Encoding and Streaming of 360-degree Video for Virtual Reality Applications. *Proceedings - 2016 IEEE International Symposium on Multimedia, ISM 2016*. https://doi.org/10.1109/ISM.2016.143

52. Girish Kulkarni, Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C Berg, and Tamara L Berg. 2011. Baby talk: Understanding and generating image descriptions. *PROCEEDINGS OF THE 24TH CVPR*. Retrieved September 21, 2016 from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.232.7714

53. Gierad Laput, Chouchang Yang, Robert Xiao, Alanson Sample, and Chris Harrison. 2015. EM-Sense. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST '15*, 157–166. https://doi.org/10.1145/2807442.2807481

54. Max Leonard. 2015. Forecaster: our experimental object-based weather forecast. Retrieved from https://www.bbc.co.uk/rd/blog/2015-11-forecaster-our-experimental-object-based-weather-forecast

55. Sabine Lösch, Thomas Willomitzer, and Gabriele Anderst-Kotsis. 2016. Snapscreen: Linking Traditional TV and the Internet. In *Proceedings of the 14th International Conference on Advances in Mobile Computing and Multi Media* (MoMM '16), 244–249. https://doi.org/10.1145/3007120.3007139

56. S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan. 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. In *2011 IEEE International Conference on Rehabilitation Robotics*, 1–7. https://doi.org/10.1109/ICORR.2011.5975346

57. Steve Mann, Jason Nolan, and Barry Wellman. 2002. Sousveillance: Inventing and Using Wearable Computing Devices for Data Collection in Surveillance Environments. *Surveillance & Society* 1, 3: 331–355. https://doi.org/10.1350/ijep.2011.15.2.373

58. Gabriela Marcu, Anind K. Dey, and Sara Kiesler. 2012. Parent-driven use of wearable cameras for autism support. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*: 401. https://doi.org/10.1145/2370216.2370277

59. Nicolai Marquardt, Ricardo Jota, Saul Greenberg, and Joaquim a. Jorge. 2011. The continuous interaction space: Interaction techniques unifying touch and gesture on and above a digital surface. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6948 LNCS: 461–476. https://doi.org/10.1007/978-3-642-23765-2_32

60. Britta Meixner, Maxine Glancy, Matt Rogers, Caroline Ward, Thomas Röggla, and Pablo Cesar. 2017. Multi-Screen Director. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video - TVX '17 Adjunct*, 57–62. https://doi.org/10.1145/3084289.3089924

61. Margaret Mitchell, Xufeng Han, Jesse Dodge, Alyssa Mensch, Amit Goyal, Alex Berg, Kota Yamaguchi, Tamara Berg, Karl Stratos, and Hal Daumé III. 2012. Midge: generating image descriptions from computer vision detections. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 747–756.

62. Norma Möllers and Jens Hälterlein. 2013. Privacy issues in public discourse: the case of "smart" CCTV in Germany. *Innovation: The European Journal of Social Science Research* 26, 1–2: 57–70. https://doi.org/10.1080/13511610.2013.723396

63. Simon Moncrieff, Svetha Venkatesh, and Geoff A.W. West. 2009. Dynamic Privacy in Public Surveillance. *Computer* 42, 9: 22–28. https://doi.org/10.1109/MC.2009.282

64. Kyle Montague, Daniel Jackson, Tobias Brühwiler, Tom Bartindale, Gerard Wilkinson, Patrick Olivier, Otmar Hilliges, and Thomas Ploetz. 2017. Prototyping Ubiquitous Imaging Surfaces. *Proc. ACM Conf on Designing Interactive Systems (DIS)*: 203–207. https://doi.org/10.1145/3064663.3064688

65. Meredith Ringel Morris and Meredith Ringel. 2012. Web on the wall. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces - ITS '12*, 95. https://doi.org/10.1145/2396636.2396651

66. Newcastle University. OpenMovement. Retrieved from https://github.com/digitalinteraction/openmovement

67. James Nicholson, Mark Huber, Daniel Jackson, Patrick Olivier, James Nicholson, Mark Huber, Daniel Jackson, and Patrick Olivier. 2014. Panopticon as an eLearning support search tool. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 1221–1224. https://doi.org/10.1145/2556288.2557082

68. Marianna Obrist, Pablo Cesar, David Geerts, Tom Bartindale, and Elizabeth F. Churchill. 2015. Online video and interactive TV experiences. *interactions* 22, 5: 32–37. https://doi.org/10.1145/2799629

69. Marianna Obrist, Pablo Cesar, David Geerts, Tom Bartindale, and Elizabeth F. Churchill. 2015. Online video and interactive TV experiences. *interactions* 22, 5: 32–37. https://doi.org/10.1145/2799629

70. Ofcom. 2018. *The Communications Market 2018*. Retrieved January 15, 2019 from https://www.ofcom.org.uk/research-and-data/multi-sector-research/cmr/cmr-2018

71. Office for National Statistics. 2018. Internet access – households and individuals, Great Britain. Retrieved September 6, 2018 from https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2018

72. Patrick Olivier, Guangyou Xu, Andrew Monk, and Jesse Hoey. 2009. Ambient kitchen. In *Proceedings of the 2nd International Conference on PErvsive Technologies Related to Assistive Environments - PETRA '09*, 1–7. https://doi.org/10.1145/1579114.1579161

73. Galen Panger and Galen. 2012. Kinect in the kitchen. In *Proceedings of the 2012*

ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12, 1985. https://doi.org/10.1145/2212776.2223740

74. Cuong Pham, Clare Hooper, Stephen Lindsay, Dan Jackson, John Shearer, Jurgen Wagner, Cassim Ladha, Karim Ladha, Thomas Plötz, and Patrick Olivier. 2012. The ambient kitchen: a pervasive sensing environment for situated services. Retrieved December 14, 2015 from http://eprints.lincoln.ac.uk/15375/1/DIS2012demo.pdf

75. Cuong Pham, Daniel Jackson, Johannes Schoening, Tom Bartindale, Thomas Ploetz, and Patrick Olivier. 2013. FoodBoard. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*, 749. https://doi.org/10.1145/2493432.2493522

76. Cuong Pham and Patrick Olivier. 2009. Slice&Dice: Recognizing food preparation activities using embedded accelerometers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 34–43. https://doi.org/10.1007/978-3-642-05408-2_4

77. David Pierce. 2015. Amazon's X-Ray Shows Movie Info Whenever You Hit Pause | WIRED. *Wired*. Retrieved September 5, 2018 from https://www.wired.com/2015/04/amazon-xray-fire-tv/

78. Anne Preston, Madeline Balaam, Paul Seedhouse, Salla Kurhila, Lari Kotilainen, Ashur Rafiev, Daniel Jackson, and Patrick Olivier. 2015. Can a kitchen teach languages? Linking theory and practice in the design of context-aware language learning environments. *Smart Learning Environments* 2, 1: 9. https://doi.org/10.1186/s40561-015-0016-9

79. Aare Puussaar, Adrian K. Clear, and Peter Wright. 2017. Enhancing Personal Informatics Through Social Sensemaking. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 6936–6942. https://doi.org/10.1145/3025453.3025804

80. Aare Puussaar, Ian G. Johnson, Kyle Montague, Philip James, and Peter Wright. 2018. Making Open Data Work for Civic Advocacy. *CSCW* 2, November.

https://doi.org/https://doi.org/10.1145/3274412

81. Guy Schofield, Tom Bartindale, and Peter Wright. 2015. Bootlegger. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 767–776. https://doi.org/10.1145/2702123.2702229

82. Paul Seedhouse, Anne Preston, Patrick Olivier, Dan Jackson, Philip Heslop, Thomas Plötz, Madeline Balaam, and Saandia Ali. 2013. The French Digital Kitchen: implementing task-based language teaching beyond the classroom. *International Journal of Computer-Assisted Language Learning and Teaching* 3, 1: 50–72. https://doi.org/10.4018/ijcallt.2013010104

83. Chris Shallue. Research Blog: Show and Tell: image captioning open sourced in TensorFlow. Retrieved from https://research.googleblog.com/2016/09/show-and-tell-image-captioning-open.html

84. Matthew Shotton, Frank Melchior, Michael Evans, Tony Churnside, Matthew Brooks, and Mike Armstrong. Object-based broadcasting - curation, responsiveness and user experience - BBC R&D. Retrieved March 20, 2016 from http://www.bbc.co.uk/rd/publications/whitepaper285

85. Sky Sports. 2016. Hawkeye inventor trials new app. Retrieved from http://www.skysports.com/cricket/news/12123/10260741/hawkeye-inventor-trials-new-app

86. Thinkbox. 2016. *New figures put TV viewing in perspective*. Retrieved August 1, 2018 from https://www.thinkbox.tv/News-and-opinion/Newsroom/10032016-New-figures-put-TV-viewing-in-perspective

87. Edison Thomaz, Aman Parnami, Jonathan Bidwell, Irfan Essa, and Gregory D. Abowd. 2013. Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing - UbiComp '13*: 739. https://doi.org/10.1145/2493432.2493509

88. TIME. Fewer People Than Ever Are Watching TV. Retrieved from http://time.com/3615387/tv-viewership-declining-nielsen/

89. John Titlow. 2016. YouTube is using AI to police copyright—to the tune of $2 billion in payouts. *Fast Company*. Retrieved August 1, 2018 from https://www.fastcompany.com/4013603/youtube-is-using-ai-to-police-copyright-to-the-tune-of-2-billion-in-payouts

90. Manfred Tscheligi, Boris de Ruyter, Panos Markopoulus, Reiner Wichert, Thomas Mirlacher, Alexander Meschterjakov, and Wolfgang Reitberger (eds.). 2009. *Ambient Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-05408-2

91. Marian F. Ursu, Ian C. Kegel, Doug Williams, Maureen Thomas, Harald Mayer, Vilmos Zsombori, Mika L. Tuomola, Henrik Larsson, and John Wyver. 2008. ShapeShifting TV: interactive screen media narratives. *Multimedia Systems* 14, 2: 115–132. https://doi.org/10.1007/s00530-008-0119-z

92. Marian F. Ursu, Julia Sussner, Ulf Myrestam, Nina Hall, Maureen Thomas, Ian Kegel, Doug Williams, Mika Tuomola, Inger Lindstedt, Terence Wright, Andra Leurdijk, and Vilmos Zsombori. 2008. Interactive TV narratives. *ACM Transactions on Multimedia Computing, Communications, and Applications* 4, 4: 1–39. https://doi.org/10.1145/1412196.1412198

93. Marian F. Ursu, Vilmos Zsombori, John Wyver, Lucie Conrad, Ian Kegel, and Doug Williams. 2009. Interactive documentaries. *Computers in Entertainment* 7, 3: 1. https://doi.org/10.1145/1594943.1594953

94. Roel Vertegaal and Roel Vertegaal. 2003. Attentive User Interfaces. *Commun. ACM* 46, 3: 30–33. https://doi.org/10.1145/636772.636794

95. Jingtao Wang and John Canny. 2004. FingerSense: Augmenting Expressiveness to Physical Pushing Button by Fingertip Identification. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '04), 1267–1270. https://doi.org/10.1145/985921.986040

96. Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. 2016. Interacting with Soli. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*, 851–860.

https://doi.org/10.1145/2984511.2984565

97. Chris Weller. 2017. Elon Musk on universal basic income: "It's going to be necessary." *Business Insider*. Retrieved August 1, 2018 from http://uk.businessinsider.com/elon-musk-universal-basic-income-2017-2

98. Gerard Wilkinson. Media of Things. Retrieved from https://github.com/digitalinteraction/MediaOfThings

99. Gerard Wilkinson, Dan Jackson, Andrew Garbett, Reuben Kirkham, and Kyle Montague. 2020. CryptoCam: Privacy Conscious Open Circuit Television. Retrieved May 3, 2020 from http://arxiv.org/abs/2004.08602

100. Gerard Wilkinson, Ahmed Kharrufa, Jonathan Hook, Bradley Pursglove, Gavin Wood, Hendrik Haeuser, Nils Y Hammerla, Steve Hodges, and Patrick Olivier. 2016. Expressy: Using a Wrist-worn Inertial Measurement Unit to Add Expressiveness to Touch-based Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '16*. https://doi.org/10.1145/2858036.2858223

101. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, 1083. https://doi.org/10.1145/1518701.1518866

102. Robert Xiao, Gierad Laput, Yang Zhang, and Chris Harrison. 2017. Deus EM Machina. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4000–4008. https://doi.org/10.1145/3025453.3025828

103. Robert Xiao, Julia Schwarz, and Chris Harrison. 2015. Estimating 3D Finger Angle on Commodity Touchscreens. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces - ITS '15*, 47–50. https://doi.org/10.1145/2817721.2817737

104. Yahoo Finance. Cisco's Set-Top Box Business Continues to Decline. Retrieved from http://finance.yahoo.com/news/cisco-set-top-box-business-170635200.html

105. Guohui Zhang, Ryan Avery, and Yinhai Wang. 2007. Video-Based Vehicle Detection and Classification System for Real-Time Traffic Data Collection Using Uncalibrated Video Cameras. *Transportation Research Record: Journal of the Transportation Research Board* 1993: 138–147. https://doi.org/10.3141/1993-19

106. Jamie Zigelbaum, Alan Browning, Daniel Leithinger, Olivier Bau, and Hiroshi Ishii. 2010. g-stalt. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction - TEI '10*, 261. https://doi.org/10.1145/1709886.1709939

107. Vilmos Zsombori, Michael Frantzis, Rodrigo Laiola Guimaraes, Marian Florin Ursu, Pablo Cesar, Ian Kegel, Roland Craigie, and Dick C.A. Bulterman. 2011. Automatic generation of video narratives from shared UGC. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia - HT '11*, 325. https://doi.org/10.1145/1995966.1996009

108. *A comparison of international television advertising markets*. Retrieved August 1, 2018 from https://www.ofcom.org.uk/__data/assets/pdf_file/0027/67905/attach2.pdf

109. EBU Technology & Innovation - Growing network of metadata developers makes MDN 2017 the biggest yet. Retrieved August 16, 2017 from https://tech.ebu.ch/news/2017/07/growing-network-of-metadata-developers-makes-mdn-2017-the-biggest-yet

110. ORPHEUS - BBC R&D. Retrieved April 20, 2017 from http://www.bbc.co.uk/rd/projects/orpheus

111. AI in Production. *BBC R&D*. Retrieved September 28, 2018 from https://www.bbc.co.uk/rd/projects/ai-production

112. Storyarc. *BBC R&D*. Retrieved September 28, 2018 from https://www.bbc.co.uk/rd/projects/storyarc

113. IP Studio. *BBC R&D*. Retrieved August 1, 2018 from https://www.bbc.co.uk/rd/projects/ip-studio

114. Victory by Marginal Gains: Team Sky, Data, and the Tour De France. *Harvard Business School*. Retrieved August 1, 2018 from https://rctom.hbs.org/submission/victory-by-marginal-gains-team-sky-data-and-the-tour-de-france/

115. PornHub Video Tags.

116. 2014. How young viewers are abandoning television. *Telegraph*. Retrieved August 1, 2018 from https://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/media/11146439/How-young-viewers-are-abandoning-television.html

117. 2016. Introducing Replay Highlights, Periscope Tweet Embeds and Live Autoplay. *Medium*. Retrieved August 1, 2018 from https://medium.com/periscope/medium-com-periscope-tweet-embeds-replay-highlights-live-autoplay-c01629ee8910

118. 2016. This is what happens when an AI-written screenplay is made into a film. *The Guardian*. Retrieved August 1, 2018 from https://www.theguardian.com/technology/2016/jun/10/artificial-intelligence-screenplay-sunspring-silicon-valley-thomas-middleditch-ai

Appendix: Contributed Projects Source

Media of Things, Meta-data Capture Platform – Capture and storage of production meta-data using IoT sensors.
https://github.com/digitalinteraction/MediaofThings

Expressy, Touch interaction augmentation using a wrist-worn IMU – adding expressiveness to touch.
https://github.com/digitalinteraction/Expressy

Expressive Things, Cross-interface Smart Home control – focussed at the nexus of interfaces on Immediacy, Context and Control.
https://github.com/digitalinteraction/ExpressiveThings

CryptoCam, OCTV Prototype – an open approach to CCTV camera footage sharing.
https://github.com/digitalinteraction/CryptoCam

MPD-L2OD, Utility to convert MPEG-DASH manifests from Live to On-Demand for later consumption.
https://github.com/GerryWilko/MPD-L2OD

OpenMovement AxLE App, Open-Source Fitness tracker, Bluetooth Comms application with data sync service.
https://github.com/digitalinteraction/OpenMovement-AxLE-App

OpenMovement AxLE Comms, Library for Communicating with AxLE Device.
https://github.com/digitalinteraction/OpenMovement-AxLE-Comms

OpenMovement AxLE Firmware Updater, iOS Firmware Updater App.
https://github.com/digitalinteraction/OpenMovement-AxLE-Updater-iOS

OpenMovement, WAX9 packet unpacker for C# and Swift.
https://github.com/digitalinteraction/openmovement