

CREATING OBJECT-BASED EXPERIENCES IN THE REAL WORLD

Michael Evans, Tristan Ferne, Zillah Watson, Frank Melchior, Matthew Brooks, Phil Stenton and Ian Forrester

BBC Research and Development, UK

ABSTRACT

The move towards end-to-end IP between media producers and audiences will make new broadcasting systems vastly more agnostic to data formats and to diverse sets of consumption and production devices. In this world, object-based media becomes increasingly important; delivering efficiencies in the production chain, enabling the creation of new experiences that will continue to engage the audience and giving us the ability to adapt our media to new platforms, services and devices.

This paper describes a series of practical case studies of our work in object-based user experiences since 2014. These projects encompass speech audio, on-line news and enhanced drama. In each case, we are working with production teams to develop systems, tools and algorithms for an object-based world: these technologies and techniques enable its creation (often using traditional linear media assets) and post-production; transforming user experience for audiences and production.

INTRODUCTION

In 2014 BBC R&D presented an IBC paper on object-based broadcasting [1], the representation of media content by a set of individual assets together with metadata describing their relationships and associations, and the abilities to bring these back together again to make new content experiences. This work has continued to progress since those very early prototypes and proofs-of-concept. We have now created a range of object-based experiences, together with experimental tools to enable the sustainable creation of such content. Collectively, these systems have formed a valuable catalyst for building our knowledge and understanding of how producers of creative content can design and deliver these experiences. We find them to be useful, practical case studies that should enable broadcasting organisations to thrive among the new broadcasting systems evolving from end-to-end IP and ubiquitous computing.

In each of the scenarios described here, we have worked with production teams to develop systems, tools and algorithms for an object-based world: these technologies and techniques enable its creation (often using traditional linear media assets) and post-production, transforming user experiences for audiences and enhancing the craft of production professionals. In [1] we emphasised the continued importance of skilled craft in

the curation of audio and video objects, as well as the data objects that describe them and their relationships and roles in the audience experience. This included the construction of a layered curatorial model, relating richer description of content relationships to more responsive experiences. In this paper we will see this distilled into the craft and the opportunities in curating the semantics of objects, exploiting descriptive relationships between experiences and the elements comprising them. Specifically, this paper describes the following projects:

- Discourse – a text-based semantic editing system for audio production
- Atomising News – structured storylining of content to support dynamic presentation
- Squeezebox – a tool for adding prioritisation semantics to segmented linear content, to allow simple control of content duration
- StoryExplorer / StoryArc – presenting an interactive experience based on the semantics of a drama and assisting writers in their craft.
- Visual Perceptive Media – a pilot of a richly annotated set of video assets, which can be assembled into a short drama based on each viewer's current context.

DISCOURSE: SEMANTIC AUDIO EDITTING

Speech radio listenership remains high and podcasting continues to grow in popularity. Although much speech content is still broadcast live, a large proportion is pre-recorded and the experience constructed using audio editing software. Commonly, such tools represent sound using simple waveforms, allowing users to visually search and scan audio content but displaying very limited information. This approach does not scale well [2]. Efficient navigation and editing of speech is crucial to the radio production process. However, unlike text, speech audio must be navigated sequentially and does not naturally support visual search techniques [3]. Furthermore, the authoring of object-based experiences may also require the annotation of the speech audio with semantic mark-up describing various useful attributes; functionality not generally offered by waveform editors.

Semantic analysis techniques can be used to extract higher-level information from the audio, such as: whether the content is speech or music [4], where different people are speaking [5] or a transcript of what they are saying. Presenting *this* information to the user could allow them to navigate and edit audio content much more efficiently. They can also be used to create new experiences like responsive radio or variable-length programmes.

Over the last year *Discourse* has been developed; a semantic audio editing system that uses a text-based interface to enable users to navigate and edit speech using an automatically generated transcript. Development included a qualitative study of current radio production and evaluation of semantic editing. We found that current practice involves time-consuming note-taking and logging, before editing the audio based on these notes. The semantic editing system allows producers to complete this process up to twice as fast in some cases. However, the semantic system was not as efficient for short recordings. Participants commented that *Discourse* allowed them to navigate and edit the audio much faster and that the accuracy of the transcript was good enough for their purposes. Both of these results support previous findings [6, 7].

It is also important to acknowledge that editing decisions are based not only on what is said, but how it is said. This emphasises the importance of a multi-modal interface, which combines the efficiency of text-based working with being able to quickly listen to the audio.

The system—which is now in the process of becoming a BBC internal product—was designed to offer basic functionality for creating a rough edit, and to easily fit into the existing production workflow. A screenshot of the interface and numbered list of the main features are shown in Figure 1.

The effectiveness of incorporating semantics into craft processes highlights audio tools and experiences as being a core part of researching object-based broadcasting.

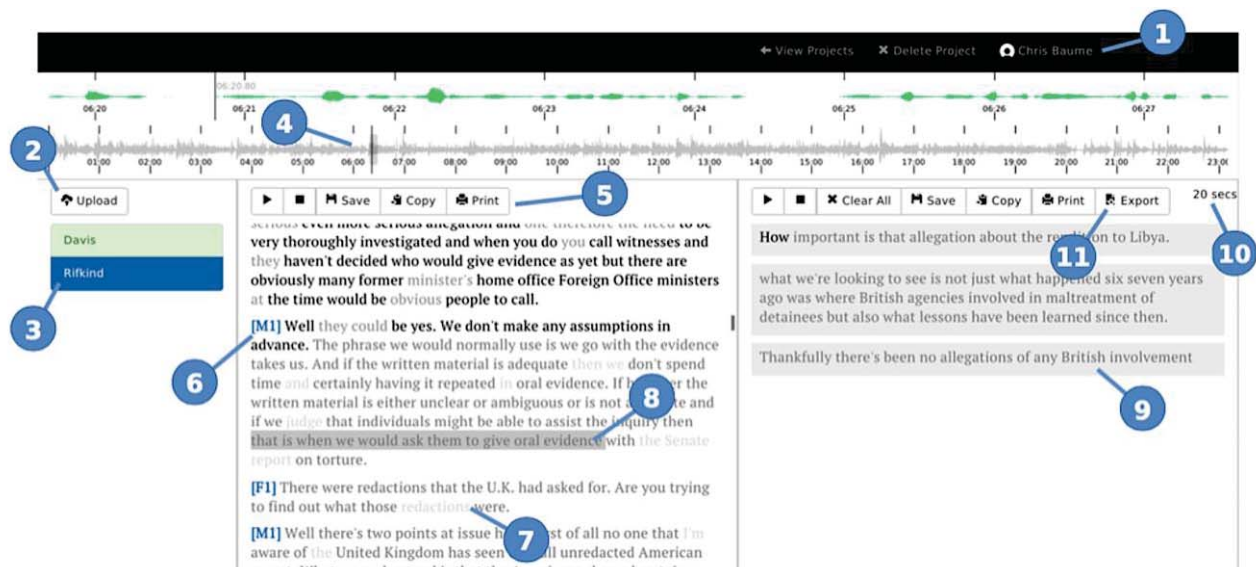


Figure 1: *Discourse* UI: (1) user accounts and projects, (2) upload audio recordings, (3) upload list, (4) waveform display of currently selected recording, (5) toolbar with playback, save, copy and print functions, (6) transcript of selected recording with speaker labelling and word editing, (7) confidence shading, (8) transcript selection with drag-and-drop editing, (9) listing and re-ordering of edits, (10) duration of edit, (11) export edit.

ATOMISING NEWS

In a similar manner to speech audio, some forms of news and journalism (regardless of platform) also lend themselves well to semantic representation and atomisation as media objects. The Storyline Ontology (<http://www.bbc.co.uk/ontologies/storyline>) developed by the BBC and others, defines storylines as being made up of events where each event can be linked to people, organisations, places and even other storylines. Although most news is complicated and subjective (Where do stories start or end? Where do they intersect? What really happened?) these structures can be used to create new media experiences.

By splitting news stories into these “atoms” of events and actors and fitting the atoms to a data model we create re-usable bits of stories that we know when and how to use, and then can put together again in different ways. These structures give us the ability to adapt our stories to different screens, platforms, contexts and experiences. News stories can be optimised for users of smartwatches, conversational interfaces, or futuristic agents using AI. Stories can also be personalised to users, which is particularly important for news,

media and broadcast organisations. Audiences use an increasing range of platforms and devices and, rather than create more and more content for each of these new platforms piecemeal as they appear, if we structure our stories into small, reusable pieces then we can efficiently re-use them again and again.

Our initial work based on these structures in news is aiming to develop a news format suitable for a younger audience and for mobile users. Through several iterations of prototyping and testing [8] we discovered that people liked having quick, skimmable summaries of stories, but they also wanted the option to be able to go deeper and get more information on the aspects that interest them.

Our prototype presents every story structured as a storyline with key events, people and places. The initial view of the story gives you a summary on one page, so you can just skim the key events. But you can expand any of the events and dig deeper — into longer-form writing, correspondent reports, social media or video. The prototype also includes pop-up definitions for the key people, organisations and places mentioned in the story. Further pilots planned for 2016 will build understanding about how journalists can write for this format.

SQUEEZEBOX

Squeezebox is a prototype production tool that explores automated video re-editing using semantic mark-up. News and other factual content is often needed at different durations for use in different programmes. Edits will be made manually, and can be time-consuming and laborious. *Squeezebox* aims to assist in the rapid re-editing of such content, allowing new durations to be instantly produced.

Currently, we target the production of captioned news story montages. Using the tool, production users add simple metadata to a collection of news stories and caption them. News story montages can then be instantly produced by selecting a target duration. Footage uploaded to *Squeezebox* is automatically analysed and segmented into individual shots. Then, rather than manually editing the footage, the user marks-up the most relevant and important portions of each shot, indicating that the rest is a candidate for being cut. She also marks up the priority of each shot, determining: how the footage behaves as the duration is reduced, which shots will be dropped first, and which ones will be preserved. Using this metadata, *Squeezebox* enables users to adjust the duration of the story using a simple slider control. The purpose-built algorithm establishes new in-and-out edit decisions per shot and in some cases drops shots entirely.

Captions are also prioritised and an appropriate number of captions are chosen to fit a story's duration, ensuring that a maximum reading rate is not exceeded. A music bed can also be selected. As we automatically transcribe any dialog found in news story videos, it's possible to use text selection to highlight phrases and instantly create "*upsounds*" – which, if chosen, automatically duck the music bed volume, restoring it when the phrase ends. User-specified idents automatically top-and-tail the montage.

Production User Research

Empirical research found that users liked the *Squeezebox* concept, with most users seeing future potential and finding the tool easy to learn and use. Satisfaction levels with the montages produced by *Squeezebox* varied: one user with experience of editing in time-critical newsroom environments appreciated the speed at which they could re-edit content,

and was happy to cede precise control to the tool. Another user who didn't normally edit news wanted tighter control than *Squeezebox* allows.

In terms of application, users indicated that it might be suitable for use by teams who don't necessarily have the knowledge or budget for high-end professional tools. It could effectively up-skill non-editors, enabling them to create finished packages without needing to involve other craftspeople. It could also allow resource-constrained teams to reduce the overheads of producing additional social media content on top of their existing output.

STORY EXPLORER

Drama on TV frequently has a complex narrative and many of us have had the experience of watching part of a long-running series whilst thinking “*Who was that?*” or “*What just happened?*”, or forgetting a crucial event an earlier series. In effect, viewers periodically crave supplementary narrative semantics. Using object-based media we developed something to help solve this problem - for people to recap, catch-up and explore the stories from TV and radio drama.

We piloted the concept with a daily radio drama, *Home Front*, in Summer 2015 on the BBC's Taster platform for audience experiments. The prototype website that we built, the *Home Front Story Explorer* (Figure 2, see also [9]), lets listeners choose the storyline with which they want to catch up, and they can then listen to that entire storyline, or highlights, or skim through it and read the parts that interest them. They can see which characters are in each scene and find out more about them. This is not intended to be a replacement for listening to the programme, but to be useful for people who want to quickly catch up or remind themselves, or who want an introduction to a drama that's new.

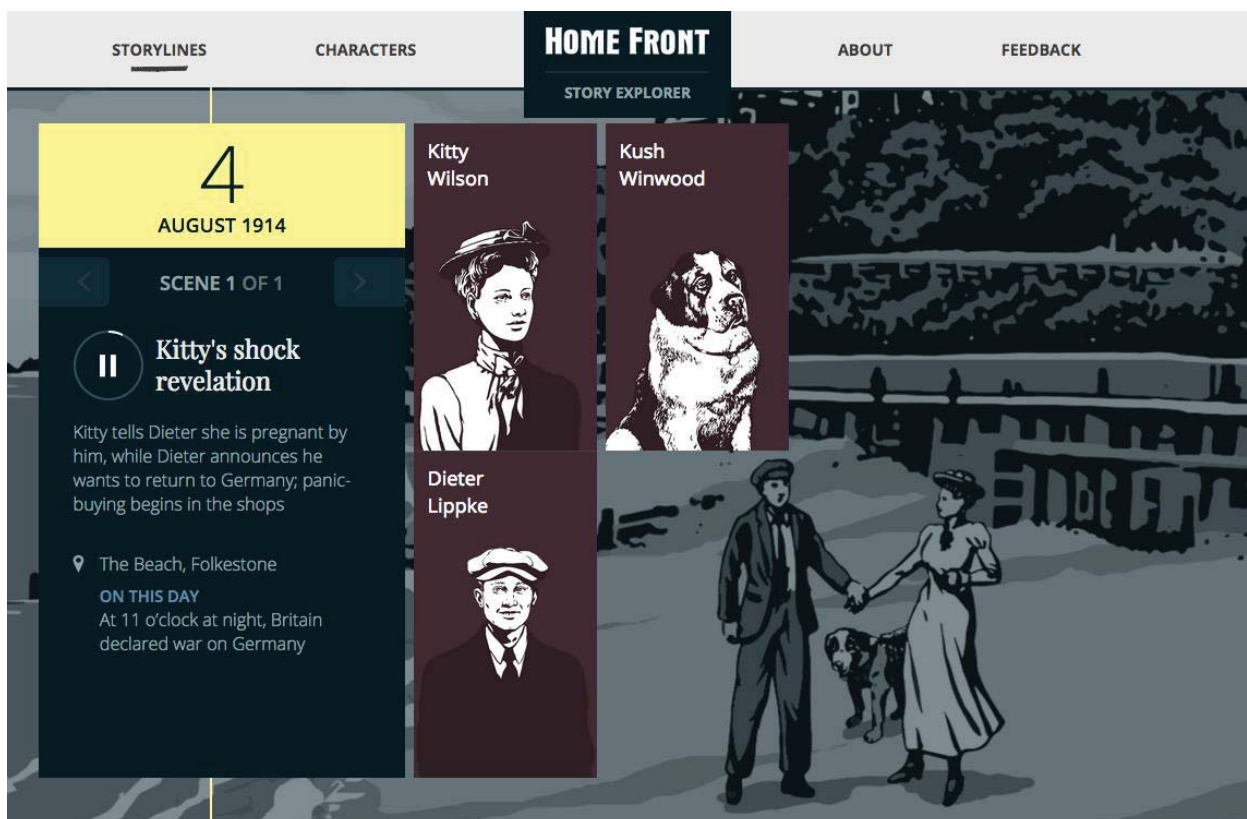


Figure 2: The Home Front Story Explorer showing a scene and the characters involved

To construct the data and media objects for this, we modelled the stories and designed the application around the stories rather than series and episodes. We used three layers to model stories:

- 1) The story world, consisting of the people (protagonists), places, and relationships;
- 2) The events that happened in this story;
- 3) The narrative: how the story is told.

Based on our research with users, the primary building blocks we wanted for the *Story Explorer* were storylines, key events and people. We took the first two series of the drama and defined the key storylines, the events or scenes in each storyline, the characters and the places involved. Then for each scene and character we snipped-out the audio from the appropriate episode and wrote a textual summary, creating a self-contained, self-described “atom” of media for every scene and character.

The pilot was very successful with Home Front’s radio audience. As expected, some were new to the drama and used the *Explorer* to get into the story. Some just wanted a recap on bits that they had missed or forgotten. Some even used it as a single-purpose radio — listening to single storylines for hours!

Our work now continues on refining and simplifying the user experience, making re-usable design patterns that can be used across radio and TV dramas [R&D blog post coming today or tomorrow] and exploring other platforms and interfaces - starting with mobile and moving on to TVs.

STORYARC

One of the challenges for the BBC in developing a new experience like the *Story Explorer* is how to scale it across the many dramas that we produce every year. Therefore, we have been working on tools that might be used to efficiently generate the structured story data and object-based media as part of existing production processes.

The Archers, 65 years old this year, is the world’s longest running soap opera. It is about a farming community and is broadcast every day on UK radio. The production has had a fairly rudimentary continuity database and an archivist for a number of years: recorded are such data as: what happened in 1967; who married who, who’s got a piano in their front room and even what the cows are called! This is used by the writers to be consistent and to give them inspiration for new storylines. We have designed a new web-based structured continuity database called *Storyarc* that replaced this old system, based upon our object-based story model. *Storyarc* is illustrated in Figure 3.

Storyarc is now in daily use by the researchers, writers and producers in *The Archers* team. It is a good example of a tool that has been integrated into existing production workflow and become useful to the team in their day-to-day work. The structured story data and object-based media produced as a by-product, can then be used to create new experiences. We deliberately designed the data model to serve both production and audience-facing needs and *Storyarc* capitalises on this. We are now looking at ways to make the authoring even easier by automatically processing existing media, including machine-parsing programme scripts to create the story data semi-automatically.

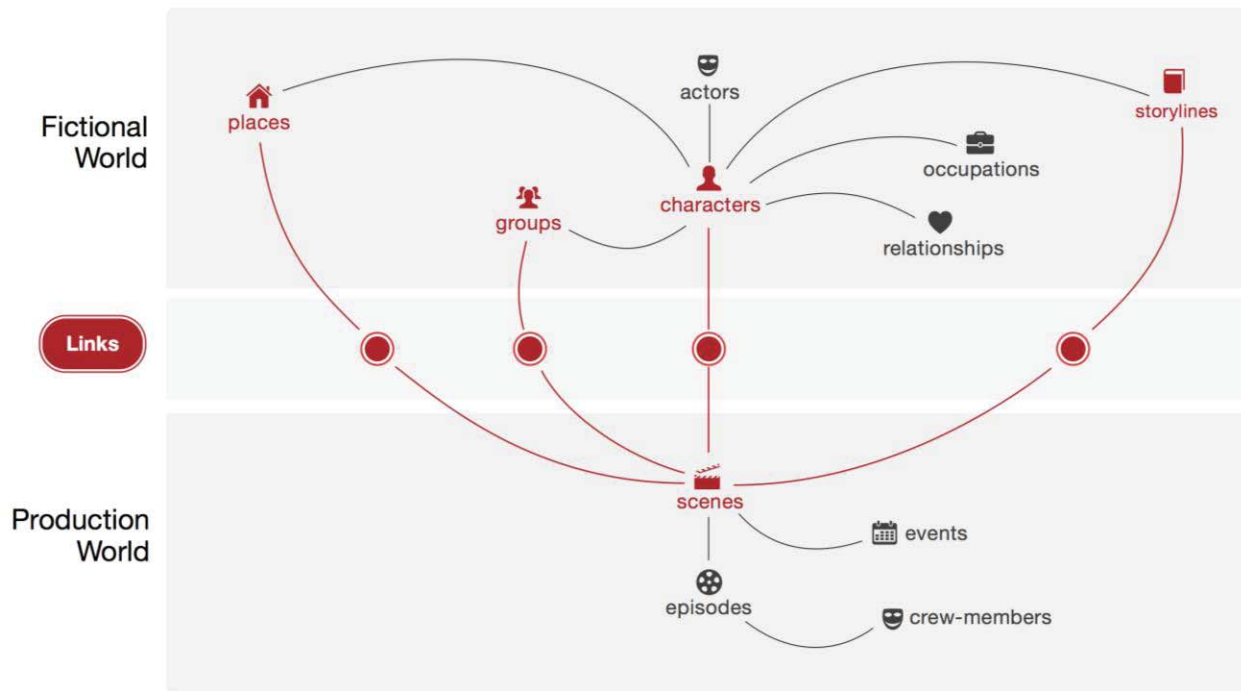


Figure 3: The *Storyarc* model showing how the data links production and fictional worlds

VISUAL PERCEPTIVE MEDIA

In this project we are researching the value that object-based content can bring, by composing each user's response to the context at the moment of delivery. If we can create content that can be composed just before or during delivery we might be able to use the semantics describing each viewer's requirements and context in order to present the most useful, enjoyable or salient experience. For example, educational content that could respond to each viewer's skill level and the challenges they encounter. Such a programme could be configured (composed) to present the right level of information and insert extra emphasis on the things *you* need to learn or see demonstrated at the time *you* need them. Perhaps the intelligibility of a particular character's dialogue requires the volume for just that character to be increased or subtitles to be included.. People who have poor sight might want the contrast of a scene to be increased to see the detail better. The main challenges are fourfold:

- How do you tag the content in a way that allows multiple re-compositions of the objects?
- What production craft is necessary to capture and produce content that might tell the same story in different ways (shorter, longer, more or less romantic...)?
- Which contextual attributes are important in order to drive composition?
- What are the requirements of the delivery engine, to support flexible composition?

As a pilot we created media assets for a simple responsive storyline, filming all the shots required to enable the flexibility composition. Audio and video were captured and stored as independent objects and a number of 'look-up tables' were constructed for colour grading. From these objects, 32 variations of a five-minute short film can be rendered algorithmically, based on the preferences and context of the audience. (Variation includes a male or female perspective, different colour grading and changes in comedic/dramatic tone.) In essence the film does not exist until it is shown; assembled just-in-time by selecting and timing individual clips, grading and soundtrack objects, from a repository.

Our ongoing research questions the challenges, opportunities and potential value of these experiences and examines how to enable a 'community of practice' in storytelling and production craft in this domain.

CONCLUSIONS

We have described several case studies of recent work in the production of object-based experiences. It is clear that curating audio, video and data assets with specific meaningful semantics to describe their roles and interrelationships, has the potential to transform user experiences for both audiences and producers. There is a range of interesting challenges that need to be investigated in order to deliver this value: Which kinds of semantic data are the most valuable in facilitating the combination of media objects? What other parts of the user experience can take advantage of object-based media—in terms of new forms of content, or more efficient production and accessible production—and how should we evaluate the practical extent of this benefit? How do we further simplify the making of object-based content, and how will the required new tools fit into established workflows? Critically, what is the balance between automation and human craft embodied in these craft processes?

REFERENCES

1. Armstrong, M., Brooks, M., Churnside, A., Evans, M., Melchior, F. and Shotton, M., 2014. Object-based broadcasting – curation, responsiveness and user experience. Proceedings of 2014 International Broadcasting Convention (IBC2014).
2. Loviscach, J., 2011. The Quintessence of a Waveform: Focus and Context for Audio Track Displays. 130th Audio Engineering Society Convention.
3. Wolfe, J.M. and Horowitz, T.S., 2004. What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience. vol.5. pp.495 to 501.
4. Panagiotakis, C. and Tziritas, G., 2005. A speech/music discriminator based on RMS and zero-crossings. IEEE Transactions on Multimedia. vol.7. pp.155 to 166.
5. Anguera Miro, X., Bozonnet, S., Evans, N., Fredouille, C., Friedland, G. and Vinyals, O., 2012. Speaker Diarization: A Review of Recent Research. IEEE Transactions on Audio, Speech, and Language Processing. vol.20. pp.356 to 370.
6. Whittaker, S. and Amento, B., 2004. Semantic Speech Editing. Proceedings of ACM SIG CHI '04, pp.527 to 534.
7. Rubin, S., Berthouzoz, F., Mysore, G.J., Li, W. and Agrawala, M., 2013. Content-based Tools for Editing Audio Stories. Proceedings of ACM UIST '13, pp.113 to 122.
8. <http://www.bbc.co.uk/rd/blog/2015-09-elastic-news-on-a-mobile>
9. <http://homefront.ch.bbc.co.uk>