Interactive Documentaries: A Golden Age

MARIAN F URSU and VILMOS ZSOMBORI
Goldsmiths, University of London, UK

JOHN WYVER and LUCIE CONRAD
Illuminations, London, UK

IAN KEGEL and DOUG WILLIAMS
BT, UK

This article is motivated by the opportunity presented by recent advancements in information and communication technology – particularly by faster broadband connections and faster digital media processing capabilities – to interactive television to extend towards and develop interactive storytelling or interactive narratives. This will give viewers the ability to shape and configure the programmes they watch, whilst watching, according to their needs and desires. Rather than consuming a predefined linear narration – represented by the traditional dramatic or factual programme – which has to address the potential audience as a whole, individuals or group of viewers can receive tailored made personal narratives. Each viewer can thus potentially become an active explorer of a narrative space rather than a receiver of a predefined narration. This article presents the production of A Golden Age, an interactive configurable documentary about the arts of the Renaissance in England, as a comprehensive illustration of the potential offered by interactive narrativity, but at the same time as a successful example for the employment of the recently developed, production- and genre-independent, ShapeShifting Media technology to the realisation of a good quality interactive narrative. The article describes the concept, the content production process, carried out from the outset with the aim of producing an interactive experience, and, finally, its authoring and delivery with the ShapeShifting Media toolkit. The focus of the presentation is on the design and implementation of the computational interactive narrative structures expressed in Narrative Structure Language (NSL), the declarative representation language underlying ShapeShifting Media. A Golden Age placed a distinct emphasis on the quality and style of each emerging individual narration, having aimed for levels at least comparable with those of (good quality) linearly compiled documentaries. NSL and the ShapeShifting Media toolkit provided the means to achieve this. A Golden Age is a production realised by Illuminations Television Ltd, London, in collaboration with Goldsmiths, University of London and BT over a period of more or less two years. Approximately 50 hours of rushes were filmed for its production. A Golden Age has already inspired the production of another similar documentary, Films of Fact, soon to be released in the public domain as an installation at the Science Museum, London, and, it is hoped, will continue to serve as inspiration for other interactive documentaries.

Categories and Subject Descriptors: H5.1 [Information Interfaces and Presentation] Multimedia Information Systems; H4.3 [Information Systems Applications]: Communications Applications; H5.2 [Information Interfaces and Presentation] User Interfaces; D2.11 [Software Engineering]: Software Architectures – Domain-specific architectures, Languages; I2.1 [Artificial Intelligence]: Applications and Expert Systems; H1.2 [Models and Principles]: User/Machine Systems – Human Factors; J.7 [Computer Applications]: Computers in Other Systems;

Additional Key Words and Phrases: Interactive Media, Interactive Television, Interactive Storytelling, Interactive Narratives, Nonlinear Narratives, Computational Narrativity, Digital Storytelling, Interactive Documentary, ShapeShifting Media, Narrative Structure Language

This research was supported by the integrated projects NM2: New Media, New Millennium (FP6 IST-004124) and TA2: Together Anytime, Together Anywhere (FP7 214793), partly funded by the European Union’s research Framework Programmes 6 and 7.

Authors' addresses: Marian F Ursu and Vilmos Zsombori, Narrative and Interactive Media Group, Department of Computing, Goldsmiths, University of London, SE14 6NW, UK; John Wyver and Lucie Conrad, Illuminations, 19-20 Rheidol Mews, Rheidol Terrace, Islington, London N1 8NU, UK; Ian Kegel and Doug Williams, Future Content Group, BT, Adastral Park, Ipswich, IP5 3RE, UK.

Permission to make digital/hard copy of part of this work for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication, and its date of appear, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.

© 2008 ACM
1. INTRODUCTION

It is widely acknowledged that traditional television, whose main function was and still is to facilitate the enjoyment of high quality moving image narrations in the comfortable and intimate space of the household, occupied a dominant position in media entertainment. Its dominance, however, has been challenged in recent years by other forms of screen media entertainment which are interactive (such as games and the web). Television responded by offering interactive features, collectively referred to as interactive television. Typically, as they are currently deployed in the consumers’ world, they

- provide interactive services, such as advanced teletext, guides, betting, trading and game-play, but in parallel with and on the side of programmes which are linear and non-interactive
- support time-shifted viewing which provides the ability to select linear programmes to watch (video on demand) and to stop and rewind through broadcast linear programmes
- offer the choice of point of view, such as selection of event or camera, but this is realised via multi-stream broadcasts, which essentially delivers in parallel a number of linear and non-interactive programmes.

For a more detailed account of such features refer to [Jensen 2005] and [Lugmayr et al. 2004], and of their relationship to narrativity to [Ursu et al. 2008a]. More recently, highly interactive spaces surrounding main narratives are provided, as places in which viewers can enhance their experience of consuming linear narrations by expressing opinions, raising issues, and, in general, discussing matters of interest relating to the topics of the narratives. At the same time, such spaces may be utilized by content creators to elicit preferences and interests from end users, in order to prepare subsequent linear narratives. However, there still is a clear separation between the main, high quality, narration, carried out through linear and non-interactive programmes, and the interactive space. In most cases, they are even physically separated: television for narration and the web (PCs and mobile devices) for interaction. Succinctly, viewers cannot interact with the professionally crafted stories whilst they are being recounted. For a more detailed account of this aspect refer to [Ursu et al. 2008b].

Despite the currently persisting separation of the interactive the spaces from their corresponding narrative spaces, there is clear evidence that narrativity and interactivity are moving towards each other and that the boundaries between them are becoming blurred [Ursu et al. 2008b] [Jenkins 2006] [Murray 2005]. This article is motivated by the opportunity presented by recent advancements in information and communication technology – particularly by faster broadband connections and faster digital media processing capabilities – to interactive television to develop interactive storytelling or interactive narratives, as a way of joining narrativity and interactivity into one single space. This will give viewers the ability to shape and configure the programmes they watch, whilst watching, according to their needs and desires. Rather than consuming a predefined linear narration – represented by the traditional dramatic or factual programme – which has to address the entire potential audience as a whole, individuals or group of viewers can receive tailored made personal narratives. Each viewer can thus potentially become an active explorer of a narrative space rather than a receiver of a predefined narration.

A rather limited number of interactive productions crafted as interactive narratives, though of quite simplistic structures, have been developed by television companies and released into the consumers’ world. The research arena in interactive moving image narratives provides richer examples but still quite limited in number, despite being more numerous than those professionally produced for consumers. However, as will be concluded in the following section, the lack of a dedicated computational framework that would provide production-independent concepts for thinking of and creating interactive narrative structures, as well as a corresponding toolkit for authoring and delivery, is conjectured as being the cause of the sparseness of such examples. Following some early but limited attempts [Agamanolis 2003][Wages 2004], ShapeShifting Media [2008][Ursu et al. 2008a; 2008b] was developed exactly to address these drawbacks: it provides a production and genre independent comprehensive language for the representation of interactive narrative structures – Narrative Structure Language (NSL) – and a corresponding generic toolkit for both authoring and delivery.
ShapeShifting Media technology has been employed in the making of a number of interactive productions in genres such as drama, documentary and news [Williams et al. 2007]. Most are still in the stage of robust and comprehensive prototypes, but one, Accidental Lovers [Tuomola et al. 2006/2007], has been successfully released to consumers, having been broadcast 12 times on Channel TV1 of the Finnish Broadcasting Company (YLE). This article focuses on another flagship ShapeShifting Media production: A Golden Age. It is an educational documentary about the arts of the Renaissance in England, configurable at viewing time through direct interaction. A Golden Age has not yet made it to the consumers’ world, but there are plans to find resources to finalise the production. However, its concept and narrative structures expressed in NSL have been adopted, adapted and extended by a more recent interactive documentary about science movies, Films of Fact [2008], set to be exhibited at the Science Museum, London, in early spring 2009. Films of Fact will be reported elsewhere.

The article is structured as follows. General considerations regarding the potential offered by interactive documentaries and an overview of notable developments related to interactive moving image narratives and interactive documentaries, presented in Section 2, provides the context for A Golden Age. Section 3 introduces the concept embodied by A Golden Age viewed from the angles of subject matter, style, and interaction and configurability. Section 4 gives an overview of the content production process, which was informed and influenced by the aim of developing an interactive experience. Section 5, the centre-point of the article, describes the development of the interactive narrative structures as supported by NSL and the ShapeShifting Media technology. Issues related to real time delivery are discussed in Section 6. A discussion of what has been achieved and of foreseeable developments brings the article to a close in Section 7.

2. CONTEXT

2.1 General Considerations on Interactive Documentaries

The documentary is a genre which lends itself straightforwardly to interaction. As people have different knowledge backgrounds, different interests and points of view, different aesthetic tastes and different pragmatic constraints whilst viewing a programme, it is quite easy to see how an interactive documentary could take them into consideration and respond to them. For example, a documentary depicting a historical period, could present a number of intertwined aspects of the respective social context, such as ethics, religion, politics, and art, or more refined ones, such as political intrigues related to a specific event, art commissioned by certain religious figures who were also art patrons, or some specific theological developments. The documentary could start by presenting an overview of the main topics, allowing the viewer to gently refine, through selection, their own topics of interest, and transforming him/her into a potential participating viewer. With the choice made, the programme’s narration could develop only or mainly along the topics of interest, but could allow further points of choice for their refinement, thus focusing more and more on the preferences of the viewer (nevertheless, points of return to the main topics may also be necessary). Choices could also be given with regards to other aspects of the narration, such as the amount of contextualisation, by allowing the request for explanations or reiteration of certain aspects already presented, depth and academic sophistication of the discourse, style of the presentation, including the ability to choose specific viewpoints and, pragmatically, overall length of the programme. These are not settings which could all be established before viewing, as, in almost all cases, they would in turn be influenced by the narration, leading thus the viewer to wanting to change them.

In general, for documentaries, interaction could be employed in selecting (influencing) aspects of the discourse such as the relevance, angle, accuracy, confidence in, and conciseness of the presented information, but also aspects related to the enjoyment and engagement value, such as pace and style of (verbal) narration and choice of illustrative material. Interactive, thus reconfigurable, documentaries also open up the possibility of

---

1 The term participating viewer is meant to point out that the viewer has now an active role in the unfolding of the narration. Synonymous terms are participant viewer, active viewer or engager. In contexts where it is clear that the viewer has an active role, the mere term viewer could be used.
refining new interpretations of the content through juxtapositions emerged through interactive viewing, but unplanned at authoring time. Furthermore, they may provide participating viewers with the ability to amend the content, thus becoming platforms for conversations, rather than voices of the experts. Whereas, indeed, watching a documentary could be a passive experience in which viewers receive a pre-prepared unchangeable narration, it could be a more rewarding experience if the viewer is allowed to interact and thus converse with the programme. Consider, for comparison, exploring a museum: following a pre-recorded tour is most probably a lesser experience compared to having a personal guide to whom we could ask questions and tell where our interests lie.

With these considerations in mind, the remainder of this section briefly surveys the related state of the art.

2.2 State of the Art

The current context of interactive TV narrativity and its relationship to ShapeShifting Media has been described in [Ursu et al. 2008b]. For convenience, some of the aspects presented there are summarised and adapted here, namely: a brief overview of the state of the art regarding the deployment of interactive TV narrative productions in the consumers’ space; a brief overview of the state of the art in research prototypes of interactive moving image narratives, focusing on documentaries, and, finally, a brief overview of two generic systems for authoring and delivery of interactive moving image narratives, which preceded the ShapeShifting Media technology, followed by a brief overview of ShapeShifting Media technology itself.

Input from active viewers, such as voting, has sometimes been used to determine the shape of programs, for example for dramatic productions, but this was done before their release, in essence following a practice common for some years in the viewing theatre context of US Studio analysis and marketing strategies. The main storyline could be chosen by viewers from a small number (normally two) of pre-prepared ones. Examples include NBC’s Law & Order: Criminal Intent [2004] and BBC’s joint episode of Casualty and Holby City [2005]. Examples of dramatic productions that respond on the fly to input from participative viewers also exist, but are very few. In BBC’s Spooks Interactive2, for example, which, notably, is delivered on the Web, the participating viewer interacts with the story-world through observation tests, logical puzzles, and reaction trials, whose scores affect the way the story unfolds. The narrative structure is very simple: two parallel paths with intermediate decision and transition points. The simplicity of the interactive structure is partly due to the constraints imposed by the current interactive television broadcast technology. Akvaario [Aquarium 2000], broadcast by YLE, is an example of an interactive production with a more complex organization of the narrative: it is automatically edited in real time from a large database of clips (approximately 5000), according to viewer interaction provided via phone. By keying a number, viewers can influence the mood of the protagonists who “react” to the aggregated input, meaning that appropriate clips are automatically chosen for delivery. However, “there is no narration, no spoken dialogue, [and] scripting was not focused on creating a story, but [on] generating a [conversational] space” [Sagas 2007]. In all these examples, the interactive narrative spaces developed are quite simplistic and, we conjecture, this is largely because of the lack of effective software to support both authoring and delivery.

Interactive moving image research productions have been described, among others, in Bushoff [2005], Reiser and Zapp [2002] and Agamanolis [2001]. They illustrate various behaviours and approaches to interactive storytelling but, for this article, two notable examples of interactive documentaries have been chosen as illustrators: Terminal Time [Mateas et al. 2000] and Vox Populi [Bocconi et al. 2005]. Terminal Time is a narrative system that produces historical documentaries which respond to the audience’s appreciation whilst they are being delivered. It focuses on the automatic generation of the narrative text, through a combination of knowledge based reasoning, planning, and natural language generation. Video clips are subsequently attached to the narration through term matching based on TF/IDF. Vox Populi generates video documentaries based on interviews about controversial topics. Via a web interface the user selects one of the possible topics and a point

---

2 http://www.bbc.co.uk/drama/spooks
of view, and the engine assembles video material from the repository to satisfy the user request. Rhetoric annotations describe the verbal information contained in the audio channel of the video segments. A specialized model encodes how these annotations are to be used to automatically generate video sequences for user specified arguments. Both *Terminal Time* and *Vox Populi* substantiate the interactive documentary format in which programmes are automatically compiled from a media pool or database of pre-recorded material on the basis of the viewers’ interaction. The interaction mode implemented by *Terminal Time* can be described as “show me more on this topic” and is available at viewing time, *during* delivery, whereas the one implemented by *Vox Populi* can be described as “show me this topic from this point of view”, but choices are available only at the outset of each viewing/delivery (once compiled, a video argument cannot be interacted with). An important conclusion to be drawn, true not only of the two examples described above, but of the great majority of the interactive narratives developed so far, is that each production was implemented by writing software in general purpose programming languages which either offer no support for moving image narrativity, or the support is very basic (e.g. overlaying video and audio). This resulted in production specific software which might be reusable with different media pools for exactly the same formats, but could not easily be adapted to amend the formats, nor could it be reused for other production formats, let alone other genres. Consequently, the concepts they implemented remained hard-coded and could not readily be transferred to other productions. Though such examples are successful in illustrating the potential of interactive TV programming, they remain mere illustrators. What is needed, we believe, certainly as a catalyst but most probably as an essential requirement, is a general paradigm offering a number of computational concepts which could be used in conceiving interactive moving image narratives, and an associated toolkit for experimentation, supporting both their authoring and delivery to end users. Notable endeavors at this end include those of Wages [2004], Agamanolis and Boye [2003], and the ShapeShifting Media paradigm [2008] [Ursu et al. 2008a; 2008b].

Though in virtual reality, Wages proposes a production independent tool for authoring what essentially is branching narratives, which entails a fixed number of points at which the decisions or actions of the engager determine the way the narrative unfolds. Interactive narratives are modelled as directed graphs, in which nodes, in turn, could be directed graphs. The recursivity of the specification is a strength of this approach, but its contribution is more or less limited to that. Agamanolis and Boye propose a generic language, Isis, for implementing interactive media, or rather, in the authors’ jargon, “responsive media”, and an associated production independent authoring system, Viper. Isis is a general purpose functional programming language tailored to multimedia prototyping, but offers an extension of primitives for implementing editing models of responsive moving image media – the editing models encapsulate the intelligence required at viewing time to compile specific narrations on the basis of engager input. Viper is implemented in Isis and provides support for media clip definition (in and out points), media clip annotation with metadata, building editing models, by exposing the primitives available in Isis, and, finally, for inspecting possible individual programmes compiled on the basis of user input and the editing model (this is a playlist viewer). The endeavour to develop a general purpose language for interactive media and an authoring system for interactive moving image productions is inspiring. Yet, from the point of view of interactive narrativity, the approach has essential limitations. Isis is a programming language offering most of the computational features of conventional languages and therefore, at this level, it is not directly relevant to interactive narrativity and also probably too difficult to be used by narrative creators. Although it offers a set of primitives for constructing editing models, they focus on editing rather than on description of narrative spaces; at this end it is more akin to SMIL [Bulterman et al. 2008] than it is to a language for interactive narratives, such as NSL [Ursu et al. 2008a]. As a final remark, Isis offers no express support for integrating interaction (cues to participating viewers and extraction of choices) in editing models. Rather, there are a number of predefined actions that may be taken when changes occur in the variables’ state, such as interrupt the current movie or end wait until it ends. Viper, on the other hand, provides an interesting interface for annotating media clips, but offers very limited support in terms of constructing narrative spaces. ShapeShifting Media with its Narrative Structure Language (NSL) and authoring and delivery software
toolkit addresses the above mentioned drawbacks and provides further features for developing interactive moving image narratives.

The ShapeShifting Media’s Narrative Structure Language (NSL) is a declarative language for specifying interactive narrative spaces and is dedicated particularly to moving image productions. It has a rigorous but non-formal semantics, which defines the way narrative structures are to be interpreted into playlists on the basis of engagers’ interaction. Any NSL specification is a narrative object. The primitive unit of composition, the atomic narrative object, is a description of a media item and, obviously, is a narrative object. NSL provides a number of aggregation structures for the recursive construction of narrative objects, including link structure, selection groups of different types, and layer structures. They are accompanied by sophisticated mechanisms for the specification of different composition features regarding the narrative objects to which they are being applied, including:

- support for the specification of both synchronous and asynchronous engager interaction; these are modelled via an inbuilt structured annotation that can be associated with any narrative object
- support for inclusion of domain ontologies
- support for annotating narrative objects, and regions of atomic narrative objects
- enabling conditions on arcs in link structures
- implicit references to narrative objects via expressions
- selection expressions in selection groups
- disambiguation and default rules (both user defined and predefined) for selection groups and for decision points in link structures
- specification of ordered sets of constraints regarding the selection and sequencing of narrative objects within selection groups; the ordering represents strength of preference
- selection groups with dynamic content, that is content provided at viewing time
- alignment conditions and synchronisation messages for narrative objects on different layers; essentially, these are constraints regarding timing of objects that are to be played concurrently
- given the recursive nature of the composition structures, alignment and synchronisation is practically possible between any narrative objects of the narrative space
- collision rules that disambiguate contradictory alignment or synchronisation conditions
- totally recursive primitive composition structures
- action points, that is annotations of points in the narrative space, possibly within atomic narrative objects, with predefined actions; an action is interpreted when the respective point is reached in the interpretation of the narrative space; an action may be a segment of code, in which case the code is executed; this is called code point
- procedural hooks; they are like action annotations, but with the effect of triggering an external procedure
- context variables which can be used in any expressions and also updated through code point annotations; predefined context variables are updated automatically
- know-how rules; they are normative statements that encapsulate some expertise regarding the compilation and editing of narrative threads; they may make reference to any aspect of the narrative as long as that aspect is captured in the associated ontology or is directly represented in NSL; if a rule is expressed within a narrative object, then it applies to any sequence of media items computed on the object’s basis to the effect of possibly amending the initially computed sequence.

NSL v1.0 is specified in [Ursu et al. 2006], but a good overview is given in [Ursu et al. 2008a]. It has a readable syntax but is not intended to be used directly by media producers; an Authoring toolkit wraps it up with easy to use graphical representations.

The production of the media content is outside the scope of the authoring environment, as it is already
supported by a number of commercially available professional tools (such as Avid and Final Cut Pro). The Authoring toolkit supports the following processes: content ingestion as media items; ontology definition; annotation of media items and segments thereof; creation of narrative structures using as many narrative canvases as needed; syntax checker and predictive text; preview for testing. Three snapshots of the authoring environment customised for annotation, narrative structure authoring and preview are presented in Figure 1.

Figure 1 Snapshots of the authoring toolkit customised for: annotation, narrative structure authoring, and preview, respectively.

Delivery of ShapeShifting Media productions is achieved via two generic components: an NSL interpreter, which reasons with narrative spaces and end-user interaction to produce fragments of playlists (in between interactions) and a content synthesiser, able to render content in real time according to specifications received in the form of fragmented playlists.

A Golden Age was realised with the ShapeShifting Media technology. The remainder of this article is dedicated to the production of A Golden Age.

3. CONCEPT

A Golden Age is a production realised by Illuminations Television Ltd, London, in collaboration with Goldsmiths, University of London and BT, with the production- and genre-independent ShapeShifting Media technology [ShapeShifting-Media 2008; Ursu et al. 2008a; 2008b]. On the face of it, A Golden Age does not appear to radically innovate in terms of the interaction mode, as, ultimately, it is a documentary which allows viewers to express their interest in the topics presented and subsequently adapts to the choices made by the viewer(s). However, the emphasis that A Golden Age places on the quality and style of the emergent narration (during each viewing) and the mechanisms employed in implementing the interaction together with the fact that it was realized using a generic authoring and delivery system, sets it apart from the other examples.

A Golden Age was conceived to explore the full range of cultural expression during the final two decades of the sixteenth century in England, which have long been recognised as a particularly rich period for the arts, with Shakespeare and Marlowe writing for the new theatre, Tallis and Byrd composing music for the court, great houses including Hardwick Hall and Longleat being constructed, Spenser and Sidney working with the forms of epic poetry and the sonnet, and more. The masterpieces of the time are set against the backgrounds of domestic and European politics, and of the rapid social change and religious conflict in these years. The ambition to cover a full range of cultural expression sets it apart from most cultural documentaries for television, which tend to focus on a single artform; the recent Simon Schama’s Power of Art (2006, BBC), for example, is exclusively about the visual arts. This distinguishing feature could partly be attributed to the interactive quality of the production – A Golden Age could be regarded as an interactive storyteller with knowledge of the cultural landscape of the time, able, for example, to automatically construct narrations on single artforms or combinations thereof, in accordance to the viewer’s interest.
A Golden Age evolved from conventional linear documentaries produced by Illuminations for UK broadcast channels including Five, BBC4 and Artsworld (now Sky Arts). Visually, and in terms of its handling of audio, A Golden Age extends the distinctive style developed in the films Gothic: Art for England 1400-1547 (2004, Five) and Encounters: The Meeting of Asia and Europe 1500-1800 (2005, Five), both of which were commissioned by Five and also released on DVD. The style of these documentaries is “classical” in television terms: straightforward, clear, focused. There is a concentration on, and a weight given to, the featured artworks, which are filmed (as often as is feasible from the original) in a comparatively unmediated manner so as to reveal their specific material qualities. The artworks are complemented by interview contributions from academics, historians and curators and by relevant generic images. An off-screen narrator provides information and context, but does not seek to impose a strong argument or a distinctively personal position. There is no on-screen presenter.

A Golden Age sets out to enhance this approach to cultural and educational programming with interaction and (re)configurability. The engager experience was intended to remain close to the familiar linear form of television, albeit with the enhancement that the engager shaped – and recognised that they were shaping – the direction and focus of the linear experience, during each viewing. The “classical”, comparatively “neutral”, visual style and overall approach were adopted as likely to be well-suited to the interactive production. There are a number of predefined themes according to which the material is structured and therefore could be explored, such as Architecture, Church, Theatre, Shakespeare, etc. At appropriate times, such keywords appear on the screen (refer to Figure 2), one at a time, and the active viewer has the ability to express a certain level of interest in them by pressing the OK button of the remote control a number of times – more for more interest. The cumulated set of interests up to that point determines the way in which the narration unfolds, but this is mediated such that the more recent the choices is, the more influence upon the unfolding story it has; conversely, this can be regarded as a decaying mechanism for the engager’s (past) choices.

![Figure 2](https://example.com/image.png)  
**Figure 2** Snapshot of the interaction interface: Architecture is the theme; two lamps lit represent a medium interest in Architecture expressed by the engager (OK button pressed twice); the side bar triangle shows how much time is left to interact. (© Wyver 2007)

The basic structure that implements this behaviour is depicted in Figure 3. There is a pool of pre-prepared media objects, essentially representing the fundamental units of narration. They are the basic building blocks with which each particular narration – i.e. during each particular viewing – is constructed, and are called here simply narrative blocks. Particular narrations – i.e. particular documentaries dynamically configured to a specific engager input – develop as sequences of narrative blocks. The sequences are compiled automatically by the narrative engine, which selects and orders different narrative blocks from the pool to both best satisfy the
engagers’ expressed interests in the themes offered and to preserve the narrative continuity and the style. Evidently, a prerequisite to achieving this is to have descriptions (metadata) of the narrative blocks and the relationships between them, from which their relevance to themes of the documentary, and therefore to the choices of the engagers, and their narrative and stylistic function could be inferred. Figure 3 only illustrates a simple thematic description which captures, for each narrative block, the themes it covers and also the extent to which each theme is dealt with (high, medium or low), or, in other words, the relevance (high, medium or low) of the narrative block to the themes it covers.

This concept poses numerous pragmatic questions, including those regarding:

- the granularity the narrative blocks: too small a size might lead to an erratic picture; too large a size might not allow for sufficient response (configuration);
- the frequency of the interaction: how often should the engager be allowed to interact? too many interaction points might disturb the TV like experience; too few might not allow sufficient scope for engager choice;
- the delay of the response: an immediate response might affect the continuity of the narration; a tardy response might not be perceived as a response at all;
- acknowledging the interaction: is it sufficient to simply adapt the programme, or should there other cues be given to the engager as acknowledgement of their interaction;
- matching content to engager choices: how should the relevance of the content be inferred to particular engager choices?
- decaying previous choices: what should the pace of decay be? a too slow decay might make the narration appear as insufficiently responsive; a too quick decay might affect the topic continuity of the narration;
- ensuring story fluency: what descriptions are necessary to provide for this automatically?
- the trade-off between responding to interaction and ensuring story fluency;

They have to be addressed when the interactive narrative structures are implemented. However, they were not regarded as essential for the content production phase. The concept proved sufficient: A Golden Age was planned to be a TV like sit-back experience, with the programme adapting in good time to the choices of the viewer, preserving the continuity and the style of the narration, and not jumping to new content as soon as a new choice is made. The pacing was conceived as measured and unhurried. The experience for the engager was to be one of looking and learning, and not one where the screen was constantly demanding attention with a frenetic
visual experience. Any individual experience was aimed to be that of a well edited linear documentary, but closer to a classic television production, like, for example, *Civilisation (1968, BBC)*, than to the rapid cuts of contemporary television drama or rock promos. Nevertheless, as a configurable documentary, *A Golden Age* was to offer the participant viewer a potentially more useful and more compelling experience than a conventional linear programme for television.

4. CONTENT PRODUCTION

The content of *A Golden Age* comprises filmed material of interviews, works of art and architecture, and performance, together with additional audio recordings of music and of a scripted narrator’s voice. The majority of the footage was created specifically for the production, although some use was made of existing archive footage from Illuminations. This reuse of material was possible because, as already mentioned, the basic visual style for *A Golden Age* was developed from Illuminations’ previous productions. It would have been far harder to integrate satisfactorily in *A Golden Age* footage filmed in other contexts with distinct approaches to framing and camera movement.

The filming was planned and organised with only an outline written treatment of *A Golden Age* in place. The parameters of the subject matter were agreed early on, and from this the most significant topics, authors and artworks were selected. This list shaped the interviews that were sought initially; for example, the production wished to cover Tudor architecture, for which the leading scholar is Paula Henderson. However, the list also included figures who were known to be general experts on the period and who were asked by the production to address specific subjects.

A similar process was undertaken with the artworks and performances: certain elements were necessary, such as the music of Byrd and the language of Shakespeare, but there was a wider range of possibilities to fulfil these, and the selection of the elements eventually filmed was often inflected by availability, cost and the interests of the performers. This partly planned, partly serendipitous approach is comparable to that which occurs in linear production, but the opportunity for the inclusion of unexpected or eccentric elements is greater with configurable/interactive productions, simply because of its wider canvas.

Filming for *A Golden Age* began in November 2005 and continued to May 2007. An extensive amount original footage was created: approximately 50 hours of rushes. This was essential for two reasons:

- to ensure a strong, consistent visual style for the production as a whole;
- to produce footage with rights (interviews, artworks and performances) cleared for the envisaged uses.

The footage was shot on PAL Standard Definition (SD, 625 lines) DV CAM with a Sony DSR 500 camera in full-height anamorphic 16:9 widescreen. Broadcast television is currently moving towards the origination of documentary material in High Definition (HD, 1080 lines), and ideally the footage for *A Golden Age* would have been produced in this format. Cost considerations, however, ruled this out, and it was recognised that the public broadband delivery networks envisaged for *A Golden Age* would not, during the lifetime of the project, be capable of delivering HD images in a form close to their original quality.

4.1 Interviews

Thirteen interviews were recorded with academics and historians to provide the core content. The interviews were filmed in a manner directly comparable to that used for linear documentaries, although their length (often more than one hour) and the breadth of topics covered was dictated by the need of the coverage: in certain instances, they needed to cover a range of general topics, whereas in other, they needed to be very focussed discussions referring to a single building or a particular play. The key difference at this stage from linear production was that a “linear interview” would usually be more focussed as the production team would have a clear idea of the eventual narrative of the film. Here, the production team was shooting to create material across a “map” of the subject territory and not just for one path through it.
4.2 Art and architecture

Filming of paintings, architecture and locations such as the rebuilt Shakespeare’s Globe theatre in London, was undertaken in parallel with the interviews, and usually by the same camera and sound operator. It was important for the production team that, as far as it was feasible, original filming of the actual canvases or buildings should be undertaken, so that *A Golden Age* could avoid the extensive use of previously shot still images. The cost of rights was again a factor in this decision, but more important was the concern to provide as direct and as unmediated an experience as possible of the objects of the discourse, such as, for example, the exterior of Hardwick Hall and its rooms, or the late Tudor paintings in the galleries at Tate Britain. Filming was undertaken at approximately 40 locations and footage was recorded of more than 400 individual buildings and works of art, providing a broad range of examples to be discovered and explored through *A Golden Age*.

4.3 Performance

Two sessions were organised for the filming of poetry and drama and of vocal music. The former was undertaken in a studio in East London with four actors, each of whom performed fragments of drama, including speeches from Shakespeare’s *As You Like It* and *Richard II*, and also poems written by, among others, Elizabeth I, Sidney and Spenser. The vocal music was shot in the church of St Stephen Walbrook with the “Renaissance Singers”, an amateur choir specializing in Renaissance music under the artistic directorship of Jan Joost van Elburg. The works recorded were secular songs and both Catholic and Protestant church music, and included madrigals by William Byrd, John Wibye, and Thomas Tallis.

4.4 Filming for Interactive Documentary

Throughout the filming, the production team aimed to produce footage that could be used in a range of different contexts and juxtapositions within the interactive form of *A Golden Age*. Concerns to ensure this during the filming at times mirrored those used during the filming of linear documentary material, such as the request to interviewees to answer questions with complete sentences and without assuming the inclusion of the question, or the need for camera movements to begin and end with frames held for a number of seconds so as to facilitate editing.

With the “classical” style mentioned above that was used throughout the production, the main difference in filming *A Golden Age* related to the length and extent of the material shot: the interviews were typically longer, reflecting the absence of a specific focus or “line” and the intent to cover more ground; the filming of buildings and artworks sometimes involved taking more shots than in a linear production, so as to ensure greater choice for both the editor and the engager.

5. AUTHORING THE INTERACTIVE NARRATIVE STRUCTURES

The interactive narrative structures were implemented using Narrative Structure Language (NSL) and the corresponding Authoring toolkit of the ShapeShifting Media technology. These are described in [Ursu et al. 2008a] and a reading of this article will significantly benefit the understanding of this article. In order to introduce the basic terms required for the description of *A Golden Age*, a very succinct overview of the ShapeShifting Media paradigm it is presented in the following subsection. The remainder of the section, then, presents the details of the implementation.

5.1 Authoring ShapeShifting Media: Generalities

A ShapeShifting Media production is an interactive screen media narrative, authored, made with recorded essence, and able to respond to input from viewers/engagers. A ShapeShifting production could lead to different
experiences depending on the engager’s interaction. A manifestation of a ShapeShifting production for one engager and during one performance or screening can be called a particular or specific narration, or, more rigorously, a narrative thread. Narrative threads are automatically compiled/edited at delivery time on the basis of the engagers’ interaction.

In ShapeShifting Media the essence is separated from the symbolic representations. Essence is ingested in a media pool as files and is represented as media items – a media item is a symbolic representation of a contiguous fragment from an ingested file, defined by an in- and out-point. The basic building block of narratives is the media item. Media items are automatically edited together at delivery/viewing time into narrative threads. Production configurability or adaptivity is achieved by (automatically) rearranging media items in narrative threads. Media items represent the atomic elements of content as far as configurability is concerned – they can be rearranged, but their content is fixed.

Media items have structured descriptions associated in the form of metadata, relating to both low-level features, such as dominant colour, and high-level features, such as the theme addressed by the item. Any media item that is included in a narrative/production becomes an atomic narrative object – the atomic narrative object inherits all the properties and structured descriptions of its corresponding media item. The collection of all the atomic narrative objects for a given production represents a part of the production’s narrative space. For each production, the terminology for describing high level features is grounded in a formally represented ontology, which, too, is part of the narrative space. The narrative space has another component, its main component, namely the production’s narrative structure, which specifies the flow of the narration with reference to the possible engager interactions. In other words, the narrative structure specifies the order in which the media items are to be edited into narrative threads in the context of any possible performance.

Narrative structures are expressed in Narrative Structure Language (NSL). Simplistically, NSL supports the recursive aggregation of media items into (structured) narrative objects and provides means for specifying the order in which the narrative objects should be ‘played’ on the basis of the engagers’ interaction. Examples of structures available in NSL are the link-structure, layer-structure and a number of selection-group-structures. The NSL aggregation structures are accompanied by sophisticated mechanisms for the specification of different composition features, listed in brief in the State of the Art section, and are fully recursive, in that they can be combined in any order and to any level of depth.

Narrative spaces are authored. In ShapeShifting Media, NSL is accompanied by an Authoring system which provides dedicated tools for the creation of narrative spaces.

Narrative threads are automatically compiled by a Narrative Engine. For each engager, this is accomplished via an iterative process that happens, loosely speaking, either between two consecutive interactions, or after an elapsed period of time. Each iteration produces a fragment of the narrative thread. The sequence of all the fragments, during one viewing, represents the overall narrative thread. The narrative thread fragments reach the engagers through specific Client Applications, which display their content, capture (new) interaction from engagers, and send them back to the Narrative Engine. Timing is crucial in the iterative communication between the Narrative Engine and the Client Applications: new narrative thread fragments have to be produced and delivered sufficiently early such that they reach the Client Applications before the previous fragments were completely played.

5.2 Editing: Creating the Media Items

As the media items are the basic (atomic) elements of content with which the narrative space is conceived, they are prepared – edited – outside and then ingested into the Authoring System. For A Golden Age, editing began with transcriptions of each of the 13 interviews. A “paper edit” of the transcriptions was first produced, striking
out superfluous material and sections in which the interviewee failed to put across a point successfully. Significantly less material was discarded at this stage than in a linear production, with perhaps 2/3 of the recorded material being regarded as useful for inclusion; in a linear production probably only 1/8 of an interview would be retained at this stage. The transcript was further divided up into paragraph sections, each of which ideally related a particular event or made a single important point. These were the foundations of the media items.

The interviews were then edited into media items using a conventional Avid DV ExpressPro toolkit. First, a “sync cut” (from synchronous sound) was made of just the interviewee’s words, and this was refined to remove “ums”, “ahs” and hesitations. Appropriate images were then added, both to cover the jump-cuts within the sync and visually to enhance and extend the impact of the words. For some elements, music or additional location sound (bird noises on garden exteriors, for example) was also added. More than 500 media objects were created for the production, and they were typically between 20 and 90 seconds in length. Each media items features only one contributor, who is seen in vision at least once, and their words may be accompanied by anything from 3 to 20 further shots. Each final edited element was a media object ready to be used in the construction of the narrative space; the completed media items were exported from Avid and imported into the Authoring Tools.

At a pragmatic level, issues regarding file sizes and types, constrained by different software components, but particularly by the capabilities for transcoding and streaming of the Delivery System, had to be resolved. The maximum bitrate that the Delivery System could reliably handle was 3000kbs for the WMV (Windows Media Video) format. The size of a one minute media item was, therefore, roughly 20Mb. These files have the original widescreen 16:9 format ratio of 1024 x 576 pixels.

Editing was a process which happened concurrently with the authoring of the narrative space. Whilst engaging with editing the material, questions such as “What if a separate music track will have to be used over this track?” and “What if, stylistically, this type of cut will clash with what ends up coming next?” could not be answered until the media items were used in the narrative and the narration tested. Iterations were necessary, when certain media items had to be redefined and reconstructed.

Indeed, editing and authoring the narrative space are intertwined and highly iterative processes. The form of media items is informed by the form of the individual narrations in which they may be used and, conversely, the form of the individual narrations emerges from the form of the constituent media items.

5.3 Themes

The material for A Golden Age was structured on the basis of themes. Themes are represented either as single words, such as “Architecture”, “Church”, “Court”, “Drama”, “Humanism”, “Legacy”, “Painting”, “Portraits”, “Shakespeare”, or as short phrases, such as “Elizabeth as Queen”, “Elizabeth as Image”, “Christopher Marlow” and “Popular Culture”, and are a mix of names, artforms and concepts. They provide the bases for both organising (describing) the content and eliciting engager interest. The terms are not mutually exclusive; for example, some content which is about “Entertainment” could also be about “Politics”. Some terms are included in others; for example, content about “Portraits” is also about “Painting”. In the actual implementation, these themes are used directly as keywords both in tagging the narrative blocks (refer to Section 3) and as cues to for extracting engager interest.

5.4 Narrative Arcs

The initial conception for authoring the narrative of A Golden Age was that the media items themselves, or the atomic narrative objects, would be the basic stand-alone unit of narration. That, consequently, each atomic narrative object would be classified by subject matter, level of complexity and so on, and therefore each narration would be compiled by the narrative engine as a sequence of such atomic objects. The narrative engine would be able to select the atomic narrative objects from the media pool that would be most relevant to the engager’s choices, and will also be able to apply rules which, for example, prevent unwanted repetitions and
enforce narrative continuity when sequencing them. In other words, that the granularity of the (re)configuration would be the atomic narrative object. With around 500 planned atomic narrative objects, there would be a great deal of possible permutations, therefore significantly many potential individual narrations/programmes.

During the planning and the first stages of experimentation it became apparent that this level of granularity is too refined and could lead to a “fractured” narrative – this could, for example, jump too often between subject areas, as new subjects would be introduced to the engager by each atomic narrative object, or could combine at a too fast pace different types of content (artworks and interviews), despite the existence of some rules for continuity. Though still possible in principle, and even presenting more potential, building an interactive narrative at this level of flexibility, but at the same time observing the aesthetic qualities aimed by *A Golden Age*, seemed an too ambitious aim, particularly as there was very little, if at all, prior experience in producing such interactive documentaries. In order to impose an element of structure but still offer the openness of (re)configurability, it was decided that the level of granularity of the narration would be made slightly coarser, by introducing the **narrative arc** as the unit of narration (the narrative block, as mentioned in Section 3), as a small standalone narrative structure made of atomic narrative objects.

Each arc normally contains between 3 and 6 atomic narrative objects, and there were more than 100 arcs created. The internal structure of the arcs is relatively simple and this allowed the editors to completely check them by previewing each in the Authoring System and thus assess their effectiveness, both in terms of content development and aesthetic pleasure. The majority of them were explicitly represented as NSL **link structures** employing conditional links and decisions points. For example, Figure 4 illustrates such an arc with one decision point (blue circle) in which the interest of the engager in “Architecture” is assessed: if it is high (represented here as >2), then the atomic narrative object “019P...ARGE” will be included in the viewing, otherwise it will be skipped. However, some of the arcs were also represented via other NSL structures, such as, for example, those designed to include random elements of content from a well pre-defined collection. Indeed, in *A Golden Age*, the structures of the narrative arcs were rather simple, but nothing in principle prevents such documentaries to employ more complex structures.

The introduction of the narrative arc allowed a more effective development of *A Golden Age* which could observe a sense of development for each particular narration. The narrative arc could be regarded as a miniature standalone narrative which, in itself, already ensures a certain quality of the narration, but also allows for an easier juxtaposition with other arcs, than atomic objects would allows for each other. Each arc develops a topic, an approach or some sphere of interest, which are more general categories than the themes. Consequently, each arc is annotated with a number of themes/keywords. Topics that overlap or embrace several subjects can be realised more coherently in this way. Each arc could have relevance to a number of themes/keywords at different levels. The determination of the relevance was editorially determined by the production team and this description/metadata added manually to each arc.

**Figure 4** Example of a narrative arc represented as a link structure in NSL. Green squares represent atomic narrative objects. There is one decision point which includes the atomic narrative object “019P...ARGE” in the viewing if the interest in “Architecture” is high (>2).

In the implemented solution, each arc is annotated with a [keyword : relevance-factor] pair. The relevance factor is an integer: 1 for *some*, 2 for *average*, and 3 for *high* relevance. This proved sufficient for *A Golden Age*, but more sophisticated representations could have been used. Figure 5 shows two narrative arcs from the
pool together with the annotation for Arc002: it has some relevance to “Elizabeth as Queen” and “Music”, high relevance to “Architecture”, and average relevance to “Patronage”. The annotations are expressed as NSL editing rules associated with the arcs. In fact, each rule is an integer expression, which represents the overall relevance of the arc to any given choice made by the active viewer, which is a list of associations [keyword : interest], with interest being an integer between 1 and 3, similar to the relevance factor. For example, if the engager choice is 3 for “Architecture” and 1 for “Patronage”, then the relevance of Arc002 to this choice is 11 (11 = 1x0 + 3x3 + 2x1 + 1x0).

5.5 Developing the Main Narrative Structures

There are two aspects of the narration that are essential to the implementation (authoring) in NSL: the elicitation of the active viewers’ interest and the subsequent selection of the relevant narrative arcs. Recall that both the cues used in the elicitation mechanism and the terms used in annotating the narrative arcs are based on the same set of keywords.

5.5.1 Interaction

The cues given to the engagers, in the form of keywords appearing on screen, are introduced by the narrative arcs and are synchronised with the content: for example, if an interview discussing the political situation of the time is illustrated with a shot of a renowned house, then the term “Architecture” could appear on screen during the whole or a part of the shot (refer to Figure 2), inviting the engager to express their interest in this theme. This is expressed/implemented in NSL as temporal annotations, more specifically as triggering regions, associated with the atomic narrative objects. Each atomic narrative object can potentially have parts/regions which, if reached in the narration, are appropriate for asking the question “would you be interested in this theme/topic, and if you are, how interested are you?” Figure 6 illustrates the atomic object “901 OPENING 1” having three such regions and the way they have been annotated in the authoring tools. The “Start Time
represents the moment when the cue, expressed as a keyword, should appear on the screen and the interaction mechanism should be activated, and the “End Time” represents the moment when the cue should be removed and the interaction disabled. In *A Golden Age*, the annotations were made at the level of the atomic narrative objects, but were validated by viewing the narrative arcs. Not all the atomic narrative objects can be used as cues for engager input and certain atomic objects can be used to make more than one invitation for interaction (as the case is with the object from Figure 6).

The basic mechanism for implementing the interaction mode is **triggering regions** – **triggered selection groups**. A triggering region can determine a triggered selection group to be included and then removed from the playlist: the start of play for the triggered object coincides with the beginning of the triggering region and its end of play with the end of the region. This mechanism is used in *A Golden Age* as follows. The narrative arcs, which include triggering regions, as described above, are placed in a looping selection group. This means that narrative arcs are played continuously, according to some selection algorithm (which will be described in the following subsection). An interactive atomic narrative object (iANO) is placed on a parallel layer in a triggered selection group (refer to Figure 7). Note that both layers are selection groups, but only the one in which the interactive narrative object is placed is a triggered group.

The start of any triggering region encountered in the currently selected narrative arc generates the interactive object to become active (through the activation of the triggered group in which it is placed). This will cause the client interface in which *A Golden Age* is played to show the interaction cue and wait for engager input. The iANO is aware of the interaction cue, as the trigger that is sent from the narrative arc to it contains the keyword with which the region was annotated. If the engager interacts, then the choice is returned to the narrative engine via the iANO, and is associated with the keyword: it becomes a pair [keyword : interest].

For example, if the currently played narrative arc contains an atomic narrative object with a triggering region annotation, say, as:

```
trigger_region(start(time1), end(time2), keyword(Architecture))
```

then at time after the object started to play, the interaction selection group is triggered, resulting in the iANO being activated. The iANO receives its duration, (time2 − time1), and the keyword Architecture, and sends them to the client user interface. If the viewer expresses a level of interest in Architecture, then this choice is returned to the Narrative Engine and the iANO is terminated. In turn, this results in the interaction selection group being terminated and waiting for another trigger.

If the engager does not interact until the end of the triggering region is reached, then the iANO is simply terminated. In most cases, this is interpreted as there being no interest from the engager in the theme represented by the keyword. However, there is an ambiguity when a keyword cue appears immediately or very near after it

---

4 The structure described here is a simplified version of what has actually been used in *A Golden Age* in order to improve readability.
5 It is important to note that the NSL structures are declarative representations and are interpreted. Obviously, the interpretation of an object is not the same with its play. However, in order to improve readability, the explanations in this section will use, at times, procedural terms which might, wrongly, send the message that the interpretation happens in sync with the delivery/play. The principles are correctly described, but the reader is asked to be aware of this convention.
has already been presented, and the engager does not interact: it is not clear whether the engager has lost interest in the theme or whether they are happy with the previous choice. In the former case, the value of the level of interest associated with the corresponding keyword should be set to “no interest”. In the latter, it should be left to its initial value. The best compromise found in A Golden Age was to diminish the value of interest to its nearest possible value. However, such situations could have been avoided by introducing a few more narrative rules which would have prevented the same keyword to appear in close succession.

The selection group containing the narrative arcs is placed on the leading layer, called main-body (Figure 7), which determines the lifetime of the layer structure (main-body in parallel with interaction) to become the same as the duration of the group structure (the leading layer is the one which determines the duration of the overall layer structure). The looping through the narrative arcs is terminated through engager interaction. A special keyword, “End Soon”, appears at selected times instead of a thematic keyword. Its choice by the engager terminates the loop (but not the narration).

5.5.2 Selection

The selection mechanism is based on the theme annotation of the narrative arcs, in the form of [theme : relevance-factor], and of the interest expressed by the engager in the presented themes, in the form of [theme : level-of-interest]. Each arc addresses a number of themes with various relevance factors. After a number of interactions, the engager’s preference is a list of [theme : level-of-interest] pairs. The main problem of selection is to find a procedure which finds the narrative arc, from a set, that best matches the engager’s preference.

The engager’s preference is represented, in NSL, via context variables. Each keyword/theme has a corresponding context variable (with the same name) initialised with zero at the beginning of the viewing session. When the engager makes a choice regarding their interest in a certain theme – called a level choice – the corresponding context variable is set by the narrative engine to the level choice value. Whenever a selection of a new narrative arc is to be made, the narrative engine considers all the non-zero context variables. In fact, A Golden Age simplified matters by expressing the formula which calculates the relevance of each arc to a preference locally, in each narrative arc (e.g. refer to Figure 5). The general formula is:

\[
\text{Relevance} = \sum \text{K}_{\text{relevance}} \times \text{K}_{\text{level}}
\]

where \(\text{K}_{\text{relevance}}\) is the value representing the relevance of the arc to the keyword \(\text{k}\) and \(\text{K}_{\text{level}}\) is the value of the context variable corresponding to the keyword \(\text{k}\). At each iteration of the selection loop, the arc that has the larger Relevance (to the choices made until then) is selected for delivery/play.

The Relevance calculation in A Golden Age is multiplicative: high relevance of an arc to a theme and high engager interest in the respective theme lead to a very significant relevance factor, compared to, for example, high arc relevance but low interest, or even low arc relevance and low engager interest. For both arc-relevance and interest-level, A Golden Age uses values between 1 and 3. For example, high interest in a theme (3) and high
relevance of an arc to the respective theme (3) make a sizeable contribution to the overall relevance factor (9), whereas low interest (1) and low relevance (1) make only a minor contribution.

Choices made in the past should be less and less influential on the currently chosen content: the more recent the choice, the higher its emphasis on the selected content should be. This was expressed via a decay function (available in NSL). Each time a new choice is made regarding a level of interest in a specific theme, all the other non-zero keyword context variables, representing all the previous choices, are decremented by 1.

Experimentally we found that the relevance formula, expressed with only three integers (1, 2, 3), combined with the decay of the past choices, conveys simply and well the link between interest and the content of the arcs, and, furthermore, also preserves the coherence and continuity of the narration.

5.5.3 Top Level Story Organisation

The structure discussed so far, illustrated in Figure 7, represents the main body of the narration: it is the ongoing recount of the themes chosen by the engager. However, an entry point leading to this structure, and a concluding point, once the engagers opts out, are necessary. The top level structure of A Golden Age is illustrated in Figure 8.

The opening section introduces the interactive production. This is followed by an introduction, which is a sequence of small narrations, each gradually introducing the themes of the documentary – i.e. the choices given to the engager – in a hierarchical order, refining the specialty of the themes as the story/introduction progresses. The introduction gives the engager the possibility to define their initial preferences. The main body is initially configured to respond to these first choices and then it provides the engager opportunities to refine their preferences. Each individual narration has to be concluded in an adequate manner. This is accomplished by the “conclusions” part: a number of predefined concluding micro-narrations have been pre-prepared, which, together, cover all the possibilities of ending the documentary/story. The choice of a particular conclusion takes into account the most recently played narrative arcs, in a similar manner as the overall narration responds to a sequence of engager preferences.

This top-level organisation ensures that each individual narration is well established and flanked by an appropriate introduction and conclusion. The shortest narrations result from playing the opening, one of the possible introductions, one arc from the main body and finally the corresponding conclusion – approximately 10 minutes long. The longest narrations result from playing the opening, one introduction, all the narrative arcs and one conclusion – approximately 4 hours long.

6. DELIVERY

ShapeShifting Media is not tied to a specific delivery setup, allowing thus the possibility to implement specific delivery systems for broadcast, broadband and local PC-based setups. There are several key challenges that have to be addressed:

- many viewers/engagers must be able to access personalized stories simultaneously;
- each person must be able to interact with their personalized story with minimal latency; this is particularly relevant in the context of delivery happening over contended IP networks;
- aesthetic standards, including the high quality of the rendered dynamic audio-visual content, must be maintained when individual stories are automatically compiled at delivery time.
The general framework for delivering ShapeShifting productions (refer to Figure 9) is designed to address these requirements while accommodating the variations made within some individual components for specific productions. Two deployment configurations have been experimented with in the case of A Golden Age:

- **broadband setup**: the viewer is running a thin-client application on a PC Windows Media Center and all the other components, including the media, are located on a remote location over the network;
- **local setup**: all the components, including the media, resides on the same Windows Media Center PC; the streaming is bypassed using direct media access, and network communication is looped back through the network adapter; this setup is suitable for a standalone application published on a DVD.

The former was the initially developed solution. The latter is a more recent development.

The Delivery System is production independent, and together with the Narrative Space (the narrative structures expressed in NSL) and the Media Pool are server-side components. The Realization Engine, an NSL interpreter, compiles a dynamic stream of SMIL playlists according to the incoming interactions. It is byte-code compiled Prolog and it is wrapped in a multi-threaded C++/Qt console application that handles the timing and communication with the Interaction Manager. The Interaction Manager is in charge of mediating the viewers’ interaction between the Realization Engine and the Client Application. It provides functionality for managing the viewers’ sessions using XML-RPC. It also maintains user profile data in a relational database and, upon interaction, the Interaction Manager inserts the appropriate profile information into the client interaction before passing it on to the Realization Engine. The Content Synthesiser is a media composition engine that interprets dynamically-changing SMIL playlists generated by the Realization Engine and performs rendering operations in order to deliver a continuous audiovisual stream to the Streaming Server.

The Client Application is the representation of the production on the viewer’s device and it is production specific. In case of A Golden Age the end-user’s device is a PC running Microsoft Windows Media Center Edition. The Client Application maintains a connection with the server-side Delivery System through two independent channels: the one-way delivery of an audiovisual stream, and a two-way channel to manage the viewer’s interactions. The Client Application is initially stored server-side and it is downloaded via an Application Server (Web Server) to and runs locally on the end-user’s device. The client is initiated from within a web-browser, this loads an XHTML document which is comprised of a Media Player superimposed on a transparent Flash component (UI Manager); both are positioned absolutely within the XHTML document to ensure correct alignment. The Media Player renders the audiovisual stream, and the Flash ActionScript
component displays and handles the viewer’s interaction that is possible via a relatively simple remote control unit. Video is displayed in full screen at a fixed resolution with widescreen aspect ratio (16:9).

In case of the broadband setup, some problems were experienced with managing the synchronisation of the keyword prompts with particular events in the video. This was due to the variability of buffering and network delays when streaming video over the Internet which caused the de-synchronization of the UI Manager and the Media Player in the Client Application. A partial but working solution to this was derived: by annotating the first frame of the audiovisual stream, the initial offset between the start of the client-side rendering, the Media Player, and the start of the UI Manager can be measured and used to accordingly delay the UI Manager. Experiments show that an initial synchronization is not enough and re-synchronization should happen at least once within the lifetime of each narrative arc (i.e. once every 4 minutes). Currently, a full working solution to this is being developed.

The local setup uses direct media access and a ShapeShifting Media Technology proprietary media player. In this case, even on currently mainstream machines, the rendering quality and continuity can be assured, as the network and streaming overheads are being bypassed. High quality narrations, both in audio and video, in sync with the UI manager, were achieved for A Golden Age in tests running for the cumulated duration of the objects in the media pool (approximately 4 hours).

7. DISCUSSIONS AND CONCLUSIONS

A Golden Age demonstrates the possibility of developing a documentary form that allows configuration in real time, whilst remaining coherent and aesthetically pleasing, using the generic ShapeShifting Media Technology, and not dedicated software.

Due to the synchronisation problem between two components of the client application – the media player and the user interface manager – in the broadband setup, mentioned above, extensive viewer evaluation was not possible; experiments were carried out only with a limited number of selected viewers. However, extensive evaluations were carried out by the production team, using the Authoring System’s Preview Tool. The programme was found to generate coherent narratives which respond correctly and without discontinuities in the narration to topic choices. Whereas a partial evaluation of the aesthetic qualities could be carried out via the Preview Tool, and was successful, a full evaluation of the interactive experience was not possible, due to the reasons just mentioned. Nevertheless, overall, in terms of the final output, the ShapeShifting Media paradigm could be declared as having been successfully applied to the production of a good quality interactive configurable documentary.

The production team reported that the Narrative Structure Language (NSL) and the ShapeShifting Media software represent entirely new ways of approaching narrative construction, which don’t fit tidily into established roles in the current media production chain. However, they also acknowledged that the narrative structures of A Golden Age could not have been developed with any media tools available on the commercial market today, at least not with a comparable development effort.

The concept of A Golden Age was adopted, adapted and extended in a recent development of another interactive documentary, Films of Fact [Films-of-Fact 2008], produced by the Science Museum, London, in collaboration with Goldsmiths, University of London and BT, and using material provided by BBC and the British Film Institute. This is to be launched as three local installations at the Science Museum in early Spring 2009. The Delivery System developed as a local setup (see above) for Films of Fact is readily usable for A Golden Age. The intention is to employ it in a proper evaluation of the overall interactive narrative experience.

Te narrative structure of A Golden Age is organised on two levels. There is the micro-narrative level – the space of the narrative arcs – and the macro-narrative level – represented by the main narrative body. Refinements to these structures could be experimented with, in a search for achieving more compelling and engaging experiences, on one hand, and to devising more effective methods of authoring, on the other. Narrative arcs could, for example, have more complex structures and employ non-explicit NSL structures, which could
make use of the metadata annotations of the media items themselves. The structure of the main body of the narrative could potentially be refined to express, for example, elements of rhetoric, which could generate more compelling narrative threads.

The selection mechanism, represented mainly by the relevance formula, could be another avenue for further investigation. For example, the current selection mechanism, based on the three integers 1, 2 and 3 for expressing both arc relevance and engager interest, a decay mechanism which decrements by 1 all the previous interest choices as soon as a new choice is made, and a formula which multiplies interest with relevance, deals well with expressed choices. It covers well the situations when:

- the preferences of the viewer change in time – by responding to the most recent expressed interest, but at the same time taking into the account previous interests;
- the viewer has equally high preference in the listed themes – in which case the narration will present a lot of variety, but keeping in focus the most recent choices;
- the viewer has low or no preference for the listed topics – in which case the narration will still present a lot of variety, but through random movement from topic to topic.

However, together with the interaction mode, the formula could be further refined for more sophisticated responses. The ShapeShifting technology more or less readily supports this. Nevertheless, viewer evaluation would be required to assess the effectiveness of more refined selection mechanisms upon increasing the enjoyment value of the viewing experience.

The narrative structure of A Golden Age also ensures that the continuity of the narration is preserved, by considering not only the most recent interest choice, but also the whole decayed history of the previous choices, and by not moving to the next arc as soon as an interaction was made, but waiting until the currently playing arc was completed. Though in A Golden Age it was considered as unnecessary, for other productions, for example which have a smaller granularity, the smoothness of the narration could be enhanced by the employment of linking material in between topics, to ensure continuity in change. Such methods could be the subject of further experimentation.

The overall narrative structure of A Golden Age is such that further material could be added, without having to make any amendments to the overall narrative structure. The only requirement is that the new material would have to be authored as narrative arcs and appropriately described with [theme : relevance] tags. This format could be extremely effective for structuring archived material into meaningful and adaptive documentaries.

A Golden Age is a sizeable professionally realised production, which combines successfully responsiveness to engager choices with a high quality of the audio-visual narration. It has already inspired the production of another similar interactive documentary, and, it is hoped, will continue to serve as inspiration for others. Finally, it has also validated the ShapeShifting Media Technology which was used in its implementation.

REFERENCES


Received December 2008.