



FROM EPIC POETRY TO AI: DISCOVERING VIABLE ALGORITHMS FOR CREATING RESPONSIVE MEDIA

Mike Armstrong

BBC R&D, UK

ABSTRACT

This paper describes a search for natural ways to make responsive media, investigating the techniques used in oral performances and contrasts these with ways which start from a fixed, recorded form. It starts with the origins of storytelling and examines the way in which the epic poet constructs stories from themes and formulas as they are being performed. It points to similarities with entity–component–system techniques in games programming. It continues by looking at a form of improvised theatre from Renaissance Italy that combined oral performance with literary learning where the lack of a script avoided censorship. This was structured around scenarios for scenes that were played out through exaggerated stock characters. The computational challenge of creating this approach to storytelling in a machine involves programming interacting behaviours aiming towards a goal. Finally, the paper examines the challenge of storytelling in the form of a tour guide, where the narrative plays out in different orders in response to the movement through a museum or town and points towards the development of responsive learning experiences.

INTRODUCTION

How can we create media that truly responds to the audience, whilst still telling flowing, meaningful and coherent stories? Since the invention of the computer, much of the focus has been on using the structures of a text adventure or video game and forms of storytelling taken from modernist and post-modernist literature such as the branching narrative, as envisaged by Jorge Luis Borges in his 1941 story, *The Garden of Forking Paths*. [1] However, these forms of media can be expensive to produce because branching narratives and alternative versions require the production of many more parts of the story than any one audience member will experience. Also, the audience can often end up disappointed by these experiences because they are left wondering what they have missed out. [2] The BBC first explored the possibilities of interactive radio drama in 2001 with *The Wheel of Fortune* and the production of the three synchronised versions, following the three different characters, proved highly challenging and very time consuming [3]. Where projects have created explorable or responsive media forms they have often relied on scheduling collections of video clips that can be viewed in any order or human moderation where the interaction includes linguistic input from the audience. [4]

This piece of work has searched for established examples of responsive storytelling that may enable us to create media that both works for a mainstream audience and can be produced at-scale, on a broadcast budget. Whilst video games are definitely mainstream, their budgets and timescales of production are well beyond the reach of most broadcasters. What follows is a description of some forms of oral storytelling that provide insights into possible approaches for responsive media. These present challenges in the form of artificial intelligence (AI) problems that will need to be solved, along with the issues for the story-tellers in structuring the stories. Some of these issues were first analysed by Janet Murray over 20 years ago, and her book, *Hamlet on the Holodeck*, (recently updated) was the starting point for this investigation. Whilst the video games industry has made great progress over the past 20 years, developments in responsive storytelling have been far more modest. [5] The challenges for responsive storytelling are, in effect, core AI challenges, getting the computer to tell the story, or present elements of a pre-built story structure, in response to the needs of the audience. As with any computational task we first need to model the problem we are trying to solve and work out what functionality the code should provide. [6] We need to build good models of stories and storytelling.

FORMS OF RESPONSIVE STORYTELLING

The first problem for us in thinking about creating responsive media is that we live in a world where the story has long had a fixed, written or recorded form. Ours is a literate society with printed books and recorded performances, our very idea of storytelling is someone reading from a book. It is very difficult for us to conceive of a world without video and sound recording, let alone one without writing. This leads on to the approach of taking fixed forms of storytelling and breaking them up into fragments to enable interaction with the audience, hence the development of branching media and taking the viewpoint of different characters. The hypothesis of this paper is that this is the wrong approach and we should look to the underlying structure of the narrative to enable the interaction, not its representational form.

The earliest form of fixed storytelling came from Greek theatre, where the oral performance was dictated by a written script. This form gave rise to what is known as the Aristotilian narrative, with its idea of a complete story with a beginning, middle, and end. [7] This closed structure of the three-act play eventually lead on to the form of the modern novel and can be found in almost all one-off radio and TV dramas and documentaries. It is in contrast to the pre-literate form of epic storytelling through poetry and song which is the form recorded in the *Odyssey* and the *Iliad*. The epic form is generally episodic in nature and created as it is performed. [8]

This paper examines three examples of responsive storytelling and suggests the types of structures that could be used to recreate their responsive nature using some form of AI. The first two are the pre-literate form of epic poetry and the Renaissance theatre tradition of *commedia dell' arte*, which sits at the boundary between oral and literate modes whilst the final one is the more contemporary approach taken by a tour guide. Though tour guiding itself also has a long history, reaching as far back as ancient Greece and Rome.

EPIC POETRY – THE BARDIC TRADITION

Stories existed long before writing. Work by Julien d’Huy has examined how native stories across the world have common roots that track the paths of human migration out of Africa between 30,000 and 15,000 years ago. His work on the genealogy of the *Cosmic Hunt* tales shows how the stories changed as the people travelled across the world, incorporating the animals and experiences from their new lands. What is remarkable is how much of the original stories has been retained over thousands and thousands of years without the aid of writing. [9]

The very idea of the story was different before the invention of writing and the printing press. In an oral society, stories exist in the moment as they are being performed. Each performance is a unique event and the exact words used will vary every time, as will the details in the story. Albert Lord, built on the work of Milman Parry, in studying Yugoslav performers of epic songs in the early to mid twentieth century, a time where these oral traditions still survived. He described how for the epic singer or poet the form of the story is always changing in the performer’s mind. It is not a static entity, but a living, adaptable artistic creation. It does not have a single pure form, neither for the individual performer nor for the tradition as a whole, complete with occasional inconsistencies. These poems or songs are not written for performance but are composed during the performance in response to the needs of the patron and the setting. In a striking example demonstrating the flexibility of the form, Parry asked an experienced singer to listen to an unfamiliar song performed by a younger singer. Parry then asked the older singer if he could sing it himself. The experienced singer proceeded to retell the story with improved ornamentation and character at nearly three times the length, yet it was recognisable as the same story. [10]

The Formulas

So how does the oral poet or singer recall and recompose a story if it does not have a fixed textural form? The answer is that they understand the story as a series of high level themes and characters which outline the form of the story. The structure of the song or poem imposes a series of constraints, or “formulas” as Lord calls them which reduce the complexity of the process. This allows the storyteller to improvise around the theme of the story in a manner that has parallels with improvisation by musicians which is described by Berkowitz as “spontaneous creativity within constraints.” [11]

The metrical patterns of the poetry, the way the rhythm is created by arranging stressed and unstressed syllables, constrains the epic poet’s choice of words. This leads the poet to develop groups of words or formulas which are regularly employed to express key ideas and facilitate the construction of the story on the fly. The poet will learn formulas for the names of the characters, common actions, times and locations. The formulas often contain more words than is necessary, in order to fit the constraints of the metrical form. On top of this the poet will develop a number of substitution systems whereby the replacement of a key word provides a new meaning with minimal effort. [10]

The poet will also have a collection of themes regularly used in telling stories. These are formulaic items which convey the common events that make up the traditional tales, for example, the convening of a council, the writing and receiving letters, a wedding or a battle. These themes are combined to tell a particular story, but they are in themselves

built up from smaller parts or sub-themes which can be reused or reworked to construct other themes. So, for example, the sub-theme used to describe a person dressing for a wedding can also be used to describe dressing for battle by substituting armour for clothing the form of the clothing. The themes and sub-themes can be lengthened or shortened, as required by the addition or omission of ornamentation, so the robes or armour would be described in more or less detail. As they learn their craft the storytellers also build up their own set of characters with which to populate their stories. These characters reappear across many different stories, along with their names and behaviours.

So, the poet learns a set of parts and rules with which they can construct stories on the fly, responding to the needs of the setting. The way in which they can be constructed from flexible, re-usable elements suggests strong parallels with the entity–component–system approach used to create many computer games. This was developed for programming in games to avoid the complexities caused by the strict hierarchy of inheritance imposed by object-based programming and provides the ability to change the behaviour of an entity at runtime by adding or removing components. Also, it is worth bearing in mind that constraints, such as those imposed by the form of epic poetry and song, usually improve the performance of AI systems.

COMMEDIA DELL'ARTE - IMPROVISED THEATRE

Commedia dell'arte is a theatre tradition, originating in Italy which spread across Europe in the Renaissance through to the eighteenth century. It sits at the boundary between written theatre and oral clowning, based on a small number of exaggerated stock characters played by each of the members of the troupe. Whilst the plot varies between performances the characters played by the members of the troupe remain the same. There are superficial similarities here to the Marx Brothers films. The plays do not have a written script, which had the advantage of avoiding censorship, instead they rely on a scenario for each scene with the actors improvising the scene around the predictable behaviours of their characters [5].

The actors often memorised items from the written material of their time, including collections of stock speeches specially written for the purpose. They also rehearsed rituals of interaction between the characters, such as an argument, that could be reused in multiple settings. Whilst the subject of the argument would be different, the way in which it is played out would be well practiced between the two actors. Like the oral poets the actors would have stock responses for handling various situations which would be pulled together to navigate the narrative of each scene. The extent to which a speech or scene would be built up from memorised literature or improvisation around the character was down to the skill of the actors. [12]

So, the story is outlined as a scenario and then fleshed out through the reactions of the characters to each other by drawing on a repertoire of responses and behaviours. The complexity of the situation for the actors is reduced by playing well defined caricatures with stock responses.

Creating an AI version would involve programming the stock behaviours and interactions of the various characters and driving them towards a goal. Programming techniques like behaviour trees may be applicable. However, whilst the storage of a large repertoire of stock responses would be straightforward in a computer, providing the appropriate response in an interaction would be rather more challenging. Whilst some simple forms of chatbot can be quite engaging within their narrow scope, such as the original ELIZA [13], which simulated a Rogerian psychotherapist, more complex forms can respond awkwardly and versions that add machine learning to make them more believable have turned out badly. Microsoft had to withdraw their chatbot Tay after just one day because it had begun to use obscene and inflammatory language. [14]

THE TOUR GUIDE

If you visit a museum and take advantage of a portable audio guide you will find that the descriptions for each exhibit are static, giving you the same description regardless of what you have already seen and whether you have heard this particular description before. By contrast if you join a guided tour, a good tour guide will provide a flowing narrative around the objects as you walk round, placing each item in context of what you have already seen and how they relate to each other. The guide will introduce overarching topics such as historical periods or schools of art as they are needed and provide more detail if the audience is particularly interested.

The challenge of generating natural language for a virtual tour guide has been studied for many years by a team at the University of Edinburgh lead by Jon Oberlander through their ILEX and M-PIRO projects. Their work has built up from the description of objects rather than focusing on narrative, but the system includes the concept of a user model, tracking the trajectory between the exhibits and settings for use by children, adults and experts. The system is driven from a database of domain and object knowledge which can even be served up to the user in a number of different languages. The creation of this complete database of objects proved to be the most time-consuming part of the work and it relies upon extensive domain knowledge. [15]

The role of tour guide appeared in our investigations following a piece of work capturing 360° video and sound of Sissinghurst Castle Garden and rooms inside the castle itself. Whilst the original intention was to enable the user to navigate between parts of the garden and rooms in the house and offer up static content, such as audio and video clips that could be triggered by the user, it soon became clear that a far more compelling experience could be provided by a spoken narrative, like a radio programme, but structured in the form of a tour guide. The stories respond to your navigation. The challenge has been finding ways of structuring the narrative about the castle, the gardens, and the people who lived there, in ways that would enable parts of the stories to be told in ways that relate to your current location. For example, if you visit a part of the garden designed by Vita Sackville-West then you have to introduce her first, preferably before you reach that part of the garden, or at least on arrival. This is a real-time scheduling problem with dependencies between the different parts.

This is a more complex version of the problems we encountered in creating a radio programme of variable length. In that case, whilst the structure of the individual story lines had to be maintained, the overall sequence of items was at least fixed, and the target length was selected before the programme could be played. [16] The problem of scheduling parts of the narrative also has common features with previous work on a cookery guide. It needed a scheduling algorithm that worked out how to interleave the steps in a number of recipes and then re-arrange them if necessary in response to the user's progress. [17] The progress made in these other projects and the development of prototype tools for object-based production should help us make progress with scheduling a series of interdependent stories to make a cohesive narrative in response to user navigation.

We are now looking to share our experience with the University of Edinburgh with the intention of collaborating through our Data Science Research Partnership. One of the key advantages in creating navigable media like this, and an important motivation for this work, is that it can provide an experience where the immersion is not broken by the video or audio pausing whilst it waits for user input. The navigation can be handled in the video domain whilst the audio continues to provide the story and guide the user around the site. Indeed, immersion in the experience can even be enhanced through the ability to navigate around the experience. [5]

FURTHER WORK

The idea of providing an AI driven guide and user navigation through space requires a model which tracks the path the user has taken through the experience. From this we are aiming to develop responsive learning experiences. The model of interaction for the tour guide in a virtual museum or house and gardens involves the narrative being shaped by the user's navigation. The user navigates the spatial domain and the AI responds with a textual or audio narrative. In effect, the AI is programmed to convey a flowing narrative and the assumption that the user wishes to explore the location. Whilst the narrative may offer advice on which direction to take next, the interaction is essentially unidirectional.

The construction of a responsive learning experience is far more complex. It will require some form of dialogue, not necessarily verbal, along with a more complex model of the user's progress. In order to move beyond the form of a fixed lecture towards that of a tutor the AI needs to contain a model of the user's knowledge in the relevant topics and update that in response to the user's progress through the experience. This would enable the AI to alter the level of detail and complexity in the subject matter and guide the user to additional background information. The model of navigation through space is adaptable to the concept of navigation through a subject area, with additional mechanisms that ensure that the user has understood key concepts before moving onto subject areas that build on them.



This set of problems moves into the area of designing machines that communicate in human ways, where the language is adjusted according to the needs of the other person. The choice of words needs to be altered to avoid a misunderstanding, such as not using technical jargon. The level of detail also needs to be altered to fit the (perceived) expectation of the other person. These issues, encapsulated in Grice's conversational maxims stem from the observation that conversation is based on a cooperative principle where each participant in a conversation contributes what is needed to move the conversation towards its aims. [18] Creating a high-quality experience with meaningful dialogue will require us to move from oral storytelling to developing machines that can master the art of good conversation and cooperation.

CONCLUSIONS

Humans have told stories for many thousands of years and developed techniques for memorising and performing these stories in a flexible manner. It was the advent of writing and then print which crystallised the story into the fixed form we are familiar with today. Our world is dominated by fixed, recorded stories that have a canonical form. By contrast the oral tradition separates the story itself from the individual performance, with the result that the story takes a different form each time it is told. Some of those techniques still survive in improvised theatre and in the role of the tour guide. If we wish to create compelling stories that can respond to the audience, then we can find important lessons in the way in which skilled performers tell stories in the oral tradition.

As we try to create experiences that respond to the audience then we need to draw on lessons from the computer industry in modelling our approach to fit the nature of the experience. The three examples here suggest that we need to use approaches that include entity component systems, develop ways of modelling stock behaviours & reactions and methods for scheduling with dependencies. All three examples demonstrate the value of structures and constraints. Constraints help improve the performance of the storyteller, both in terms of recall and performance and they are also a key in achieving good performance from AI systems.

Programming machines to tell engaging, flowing stories, either through the generation of natural language as with ILEX and M-PIRO or by the scheduling of pre-recorded content, will require further improvements in our understanding of the structure of stories and how we can express the structure of stories in a way that can be created and/or scheduled by a machine. Further work is needed to reach the point where we can create experiences that include a meaningful dialogue with the audience.

DEDICATION

In memory of Jon Oberlander, 1962 - 2017, Professor of Epistemics at the University of Edinburgh, whose passion was finding ways of enabling computers to create language and adapt to people.

REFERENCES

1. Wardrip-Fruin, N. & Montfort, N. (eds.) 2003. *The New Media Reader, A User's Manual*. The MIT Press.
2. Khatchadourian, R. 2017. *The Movie with a Thousand Plotlines*. *The New Yorker*, January 30, 2017. <http://www.newyorker.com/magazine/2017/01/30/alternate-endings>
3. Greco D. & Fisher N., 2001. *god, random access, or roulette?* Hypertext Now web site, 2001. <http://www.eastgate.com/HypertextNow/archives/Fisher.html>
4. Ursu, M.F., et al, 2008. *Interactive TV narratives: Opportunities, progress, and challenges*. *ACM TOMM*, 4(4), p.25.
5. Murray, J.H. 2017. *Hamlet on the Holodeck*, updated edition. The MIT Press.
6. Gajski, D.D., Zhu, J & Dömer, R. 1997. *Essential Issues in Codesign*. In: Staunstrup J., Wolf W. (eds) Hardware/Software Co-Design: Principles and Practice. Springer, Boston, MA.
7. Aristotle, c. 335 BC. *Poetics*. The Project Gutenberg <http://www.gutenberg.org/files/6763/6763.txt>
8. Ong, W.J. 2002. *Orality and Literacy*, 30th Anniversary Edition. Routledge.
9. d'Huy, J. 2016. *The Evolution of Myths*. Scientific American, Volume 315, Number 6.
10. Lord, A.B. 2000. *The Singer of Tales*, second edition. Harvard University Press.
11. Berkowitz, A. 2010. *The Improvising Mind: Cognition and Creativity in the Musical Moment*. Oxford University Press.
12. Henke, R. 1996. *Orality and Literacy in the Commedia del'Arte and the Shakesperean Clown*. Oral Tradition, 11/2: 222-248.
13. Weizenbaum, J. 1966. *ELIZA — A Computer Program for the Study of Natural Language Communication between Man and Machine*. Communications of the Association for Computing Machinery 9: 36-45.
14. Neff, G. & Nagy, P. 2016. *Talking to Bots: Symbiotic Agency and the Case of Tay*. International Journal of Communication 10, no. Special Issue (2016): 20.
15. Isard, A, Oberlander, J, Matheson, C & Androutsopoulos, I. 2003. *Speaking the Users' Languages*. IEEE Intelligent Systems, 18, 40-45.
16. Armstrong, M., Brooks, M., Churnside, A., Evans, M., Melchior, F. & Shotton, M., 2014. *Object-Based Broadcasting - Curation, Responsiveness and User Experience*. IBC 2014.
17. Cox, J. Brooks, M. Forrester, I. & Armstrong, M. 2017 *Moving object-based media production from one-off examples to scalable workflows*. IBC 2017.
18. Grice, H.P. 1975. *Logic and conversation*. In Cole, P., and J.L. Morgan, eds. Speech Acts. New York: Academic Press, 41–58