INTERACTIVE 360° NARRATIVE FOR TV USE

Christian Fuhrhop [*], *Louay Bassbouss*[†] *and Nico Patz*[‡]

[*]Fraunhofer FOKUS, christian.fuhrhop@fokus.fraunhofer.de
[†]Fraunhofer FOKUS, louay.bassbouss@fokus.fraunhofer.de
[‡]RBB nicolas.patz@rbb-online.de

ABSTRACT

Given appropriate playout technologies, 360° video can be shown on regular SmartTVs (HbbTV), even as live video. However, limitations of TVs restrict the amount of interactivity that can be provided on top of the 360° navigation. The paper utilizes examples from the Hyper360 project to provide suggestions for the design of interactive 360° TV applications.

One of the topics addressed is the need to consider the requirements of interactive 360° videos in pre-production and during filming, providing content that will provide a viewer experience without disruptive visual and technical effects.

The paper is based on the lessons learned from the steps of building an interactive application for web and HbbTV consumption, from initial technical capability tests to the finished interactive 360° experience.

Index Terms— HbbTV, 360°, Interactive narrative

1. INTRODUCTION

The term "Interactive TV" has no clear or commonly used definition. The Wikipedia definition, for example, includes all forms of television with added data services, including online banking [1]. Some definitions see interactive TV as any TV which can be influenced by viewers, including casting shows, other define it as specific to pay-TV set-top boxes, contrasting the term to "converged TV", or distinguish between two definitions of interactive TV, one technology based, one usage based. The later one is the definition which we will use in this paper, namely "Interactive TV enables a participatory experience with content on the TV screen.". In this paper we present an interactive 360° narrative, which allows viewers to examine a fictional crime scene.

1.1. Distinction between interactive storytelling and video games

There is an overlap between interactive video games and storytelling using interactive TV. The first video based game, "Dragon's Lair" [2] can be seen as a user-influenced video narrative. The original game consisted of traditionally hand-animated video sequences, stored on a laserdisc. At various points of the video, players could press one of five buttons to influence the next sequence to be presented. On screen content was almost entirely video based. The only non-video elements were a 'lives' and a 'score' counter.

While, technically, this makes "Dragon's Lair" an interactive video experience, there are aspects that make it a "game" and not a narrative experience. Most relevant is that there's only one "correct" way to navigate through the story. While the game contains almost twice as much video content as presented in the primary storyline, all additional materials represent dead ends. The main motivation of the player is not to experience alternate narratives, but to avoid the 'wrong' video sequences and find the 'correct' path to the happy end.

For the purpose of this paper, we consider these six features important for interactive narrative TV experiences.

1. Every path should provide a satisfactory narrative - no sudden deaths
2. Content is predominantly video based - no 3D worlds
3. Viewer's motivation should be following the story - not reaching a high score
4. Casual experience - viewer should not be required to react quickly to video events
5. Interactions have an effect - no 'fake' choices
6. Media playback selections only - the viewer shouldn't have an inventory or collect

2. EXAMPLES OF INTERACTVE TV

2.1. Early examples

Probably the earliest case of interactive TV in Germany was "Mörderische Entscheidung" (Murderous Decision) [3] produced by the public broadcasters WDR and ZDF and broadcast in 1991. The story was filmed from the view of the two main protagonists and shown simultaneously on two broadcast channels, allowing the viewer to switch between the two storylines by switching between the two channels. Similar experimental approaches have been taken by other broadcasters, the most elaborate being the D-Dag project, with four different storylines shown on the first day of the year 2000 on seven different channels in Denmark synchronously. Four of the channels showed the individual

978-1-7281-1485-9/20/$31.00 ©2020 IEEE

storylines, one channel presented all four storylines as a split-screen presentation, the other two presenting background information.

Another approach was taken, for example, by RTL II in Germany in 2003 with "Jack Point Jack" [4]. While different storylines and endings were filmed, on its initial broadcast only one storyline and one ending were shown. During the programme viewers could call a premium phone number to vote for the next segment to be shown.

While there have been other pre-digital TV productions utilizing variants of these approaches (parallel broadcast and audience voting), all those shows were stand-alone experiments, seemingly for economic reasons.

With the limited number of analogue channels available, parallel broadcast of the same story on multiple channels used up resources which could have been utilized to reach a wider audience by showing content attractive to different target groups. Scenarios, which employ audience voting to select between alternatives of which only one is shown, use only one broadcast channel, but require the costly production of alternative material, which is never presented to the audience. This provides a strong economic incentive to, at least, lower the production values of parts most likely not selected and to influence the audience into selecting a preferred storyline. An extreme example here is probably "Mr. Sardonicus", where it is doubtful whether an alternative ending was ever filmed and it seems likely that the audience was strongly influenced towards the "punishment ending" and this was shown regardless of the actual audience reaction.

While audience votings have become a significant part of the current broadcast TV experience, they are now exclusively used for non-narrative formats, such as reality shows, casting shows and other competitive formats. A notable feature here is that formats are usually live and the audience votes for a person or a team, not a storyline. While the voting has an effect on the remainder of the show and future segments of the series, it does not have a significant impact on production costs, as no alternative presentations need to be pre-produced and it does not have a strong effect on the ongoing production whether candidate "a", "b" or "c" wins the show and sings the song again or is eliminated and does not take part in the following episodes.

A more suitable medium for interactive video than broadcast seemed to be DVDs. With the capability to switch between multiple viewing angles, DVDs provide the technical platform for switching between parallel storylines at ease. DVDs also allow the presentation of video segments in arbitrary order, based on viewer menu selection, making it a suitable platform for allowing the user to navigate through alternative story paths. In contrast to the use in broadcasting, where only one version is presented, DVDs allow the viewer to experience the story multiple times, selecting different branches through the story each time and viewing all alternative versions.

However, even on DVDs, and subsequently Blu-Ray discs, such interactive narrative formats are rare. There are few DVDs featuring multiple angles and in most cases, these are 'used as advertised' and show different 'angles' of the same scene, such as an Apollo 11 documentary showing the lift-off from four different positions, but not alternative storylines.

The same applies to branching stories. While these are encountered more often, their main use is the presentation of additional material, which, in most cases, have been produced for inclusion of the original movie and then been cut from release due to length and pacing reasons.

2.2. Recent developments in interactive TV

With many TVs installed in European households being some variant of smart TVs with internet connectivity, interactive video is increasingly IP-based. This allows for a larger number of alternative versions being available than fixed storage mediums like DVD or Blu-Ray would allow for, while also allowing the provision of variant content to a large number of viewers without tying up TV channels. Currently, interactive narratives are mostly provided as add-ons to existing TV or movie franchises, expanding the experience, but not being necessary to the stories told.

In 2017, Netflix produced an interactive episode "Puss in Book: Trapped in an Epic Tale" [5] as part of the "Puss in Boots" animated series, spun off from the "Shrek" films. This was used to validate the technology before producing an interactive episode of the "Black Mirror" series, named "Bandersnatch" [6]. Among the decisions to keep the episode accessible to regular viewers were the use of only two alternative story branches at every decision point, a minimum length of two minutes for every story segment, a ten second timeframe for every viewer choice, a clear visual indication when choices could be made (video layout change and textual description of choices) and a default selection for every interactive point, allowing viewers who do not use the interactive features to still experience a basic version of the episode.

3. HYPER360 INTERACTIVE VIDEO CREATION

As part of the Hyper360 project, German public broadcaster RBB created an application allowing viewers to explore a crime scene in a 360° environment, which will be described in more detail in section 5. One of the requirements was that the application would be accessible to viewers on a wide range of platforms, including web browsers and mobile devices and not solely available for TV viewing.

The Hyper360 [7] project has developed OmniConnect [8], an annotation tool for 360° video, which was utilized to create a version of the interactive 'Crime Scene'. The web-based annotation tool supports a wide range of features, including the dynamic adaptation of hotspots and visual

markers, depending on previous user behaviour and the assignment of metadata to areas and elements to support recommendations, based on ontological hierarchies.

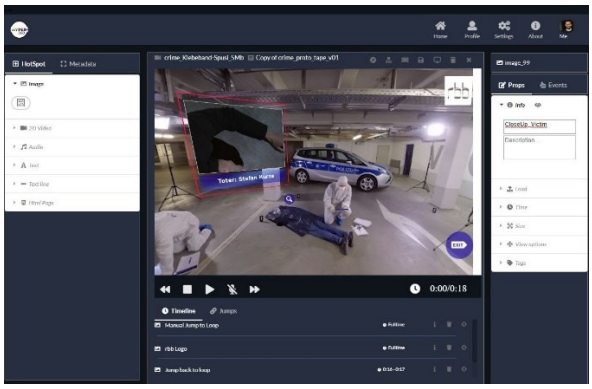

Figure 1: Web based annotation tool

The 'Crime Scene' application uses only a subset of the available features. Used are features are: linking between different videos (360° and fixed perspective videos), presentation of images, text or picture-based hotspots.

Interactive videos created with the annotation tool can be directly played on a set of Hyper360 content players, consisting of a Unity3D 360° based desktop player for Windows and Mac OSX, a web-based player implemented using HTML5 and WebGL, an iOS player written in IOS native Xcode, an Android player, which is based on the Unity3D desktop player, and a VR player for the HTC Vive. These players cover immersive platforms as well as screens with touch, mouse and remote control input devices.

3.1. Interactive video creation for TV use

Presenting and interacting with content on connected TVs requires an approach that differs from creating interactive content for other players.

All other player environments have some sort of 'pointer' function, whether activated by mouse, finger, gaze or the handheld controllers of an HTC Vive, so it is always possible to trigger an interactive element directly. Although there are televisions with pointer devices on the market, HbbTV, only mandates four directional keys on a remote control, plus 'enter' and 'back'. The same applies to the remote control for FireTV devices. Interactive TV applications need to be usable via keys on a remote control, requiring a navigation system that differs fundamentally from the other platforms.

There are also a number of technical restrictions for most TV devices, most importantly that they currently lack the necessary hardware for doing the 3D transformations required to render a specific view from a 360° equirectangular video source file. As a result, 360° videos are pre-rendered for TV use. This means that scenes are rendered for specific viewing angles, most commonly in 30° steps horizontally and vertically.

As switching between these videos, for example from a horizontal view of 30° to 60° would introduce a visually irritating jump, additional videos are rendered that show a 'pan' of the virtual camera between these fixed views. The length of such a 'pan' can be freely defined, but is in most cases a half or a full second. [9]

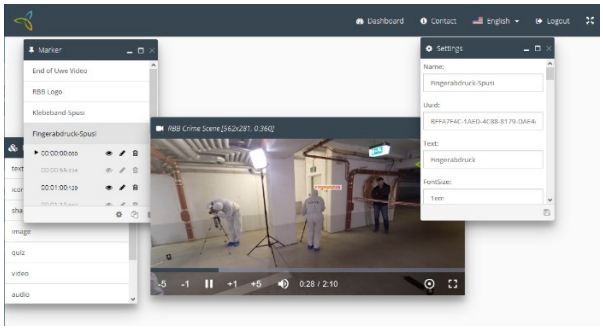

Figure 2: Tool for HbbTV interactivity editing

If the viewer presses, for example, the 'right' button at any point during the video, the current view (for example 30°) plays to the end of the current time period (for simplification we assume that the video has been rendered in one-second segments), in this case to the end of the current second, then a video segment panning to 60° is shown for the next second, from which point onwards the view from 60° will be presented.

Care needs to be taken, however, that the virtual horizontal viewing angle used in the application is at least two times the step size used in pre-rendering, so that an object or person that is on the edge of the current view can be brought roughly to the center of the screen by a key press.

Due to the different requirements of authoring for TVs compared to devices with pointers and 3D capability, Hyper360 uses a different editor for interactive TV applications. Visual elements, such as icons, pictures and videos are re-purposed for this editor, but the user interaction with the 360° video and interactive elements will be authored according to the specific requirements of TV usage.

4. APPLICATION PLANNING

Before creating the content for the "Crime Scene" application, a "technical test" prototype was built. This had three purposes. First, to get everyone involved comfortable with the software and hardware tools to build the application, second, to find out whether all features initially considered would be feasible on the target devices and, third, to reduce costs by working with test content that could be produced

cheaply and with low effort before producing the "Crime Scene" content.

For the test scene, a "walkthrough" of the RBB premises in Potsdam-Babelsberg was used, covering four locations: two outdoor locations, a radio studio and a TV studio. In each of the locations a 360° camera was set up and short (up to 3 minutes) videos of the location were recorded.

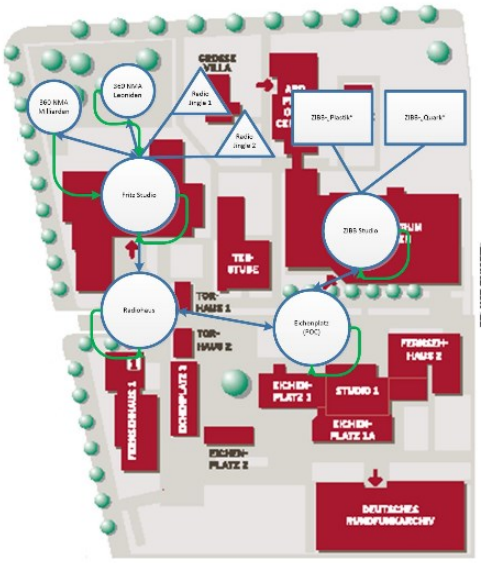

Figure 3: Relation between media elements (white with blue border) and the physical location on the RBB premises (red buildings)

Additional content used for the test scene were audio files from the radio studio, two 360° concert clips and two fixed-perspective videos of programmes produced in the TV studio.

A rudimentary map was drawn, denoting the relationship of the individual media elements (see Figure 3). Circles denote 360° videos, rectangles are fixed-perspective videos, triangles represent audio files.

Screenshots from the videos were used to position the interactive elements in the scene in a storyboard. The red line marks the desired 'initial viewing direction' for the video, when a viewer 'enters the room' from the outside. The 0° angle for the video is at the center of the picture.

Figure 4: Placeholder graphics and final look

Initially, 'placeholder' graphics were used for the interactive elements, which were later replaced by graphics specifically designed for the application.

Additionally, hotspots for rotating the view were added at the lower left and right of the screen. The player supports two modes for rotating the video with the remote control. In one mode, the directional keys directly rotate the video and interactive elements are ignored. Viewers need to press "Enter/OK" on the remote control to enter the hotspot-navigation mode, which allows selection and activation of visible hotspots. While the capability of rotating the view directly with the remote control keys provided a better user experience, viewers weren't always aware which mode they were using and how to return to directional key rotation mode. As a result, dedicated "Turn left" and "Turn right" buttons were added at the button of the screen to allow viewers to navigate and interact with the scene while staying in hotspot-navigation mode only.

Viewers can only navigate between hotspots that are in the current field of view. Allowing the navigation to and selection of offscreen hotspots caused confusion, as viewers didn't know what hotspot was currently selected and would be activated on pressing "Enter/OK". While the obvious solution would be to rotate the view to ensure the selected icon is in the field of view, viewers were no longer sure which navigation mode they were in and could no longer tell whether the view turned right because they were in direct rotation mode or because there was a hotspot they were selecting. Allowing users to only navigate between on-screen hotspots improved the usability of the application.

4.1. Technical Testing

The application was tested on various HbbTV devices and FireTV sticks.

During the tests, two technical issues were identified, which influenced the design of the "Crime Scene" application. In the TV studio scene, the original plan was to run the content from the fixed-perspective recorded TV programmes as "picture in picture" in the 360° video. One was located on a monitor in the studio, the other on a video projection wall. However, most HbbTV devices only support one active video object at a time. In most devices, this meant that the 360° video 'froze' and only the inserted video was playing. Furthermore, not all HbbTV devices showed the same behaviour. On some devices only the active video object was shown, with the other video showing a black frame, which was unsuitable for the application. To circumvent this problem, the fixed-format TV segments were shown full screen.

An unexpected technical issue was the playback of audio files in the radio studio video. While HbbTV devices should be able to play additional audio files while showing a video, some devices could not handle this. The assumption is that

the implementations on those devices use the same media object for audio and video elements, confusing the A/V control object. To avoid this potential issue, the "Crime Scene" application does not use separate audio elements.

5. CRIME SCENE" APPLICATION

The topic selected for the interactive narrative is a crime scene. While the application is designed as an independent story, it is intended to complement the fictional and real crime programmes shown on RBB by educating the viewers about the roles of individual specialists active at a crime scene.

The initial setting is an underground car park where a corpse was found. The police were notified and are performing their investigation. Viewers can look around and interact with the people and some of the objects. Green dots denote links to 360° videos, red dots are links to fixed-perspective videos and yellow rectangles are links to images.

Figure 5: "Crime scene" placement of interactive elements

Individual storyboards for all three media types denote which link will activate which content. Not all content had been produced by the time the storyboards were made. Placeholders have been used for missing videos and photos.

For all persons in the scene, there are two links to two videos, one in 360° and one fixed-perspective video. In the 360° videos, individuals give information related to the current case and what they have found out about it. For example, the ballistic expert states that a 9mm bullet has been used and where the casing has been found. Using the other link shows a video that gives background information about the work of that specialist, e.g. what a ballistics expert does and how.

Most objects also have two links. The image link leads to an 'evidence documentation photo' that shows the evidence on the crime scene. The video shows how this evidence is processed. For example, the evidence picture depicts the shoe print found on the scene, while the video shows how shoe prints are used in investigations.

As the idea of the application is to encourage exploration, all links are active and accessible to viewers. There are no dependencies; a viewer can get the information about the crime scene from the fingerprint specialist without being required to watch the video about what a fingerprint specialist does.

While the investigation develops, there are a few red herrings where the user will think about possible options. With more and more witnesses and suspects being interviewed and some more incidents found, in the end the case will be solved by the experts with the help of the guest investigator, the user.

5.1. Content Production

For the fixed-perspective videos and images, RBB used established production methods. The 360° videos were recorded using an Insta360 Pro camera.

The scene uses one fixed camera position. Actors were assigned positions around the camera, that were, roughly, equally spread around the camera. The purpose of this placement was to encourage viewers to make use of the "look around" feature of the 360° video, to provide a clearer separation of interactive hotspots and reduce the number of people visible at the same time, creating fewer motion artefacts when switching between videos.

After an initial scene with a presenter entering and explaining the situation to the viewers, actors were asked to perform 'neutral' activities with only minor body movements, to create a loopable video of approximately ten minutes, over which the interactive elements were placed.

Then each actor gave a short statement, in most cases about 30 seconds, towards the camera, presenting the findings on the crime scene. Actors stood up and removed the face protection when addressing the camera, indicating to the viewer who was currently addressing them. While not specifically asked to, other actors used this time to perform more pronounced activities, like using a tool or shifting position, taking care to be back at the 'neutral' position when the statement ended.

The individual 'statement scenes' were then linked from the 'neutral' scene, allowing viewers to navigate in the 360° view and trigger statements from all actors.

While small jumps can be observed when the view switches between the 'neutral' and the 'statement' scenes, these were generally acceptable to viewers.

5.2. Limitations

The primary challenge arose from the fact that viewers could trigger a 'specialist statement' at any time, as opposed to interactive videos that provide viewer choices only at the end of scenes. Giving this freedom to the viewer was deemed an essential aspect of the application to encourage the sense of exploration. As it was unpredictable when a viewer would trigger a 'statement scene', providing a truly seamless transition between scenes was unpractical.

Various solutions were considered to address this issue, including adapting the approach most often taken with interactive fixed-perspective videos, where transition changes are hidden by a cut. For 360° the approaches considered were a change of position, where the camera is positioned closer to the active actor, a change of perspective, where the field of view 'zooms in' on the actor, a transition effect or switching to fixed-perspective video for the statement. However, all these approaches are disruptive and have a negative effect on the immersiveness of the 360° experience, undermining the initial reason for using 360° content.

Another option was recording actors separately and compositing them into the scene during playback. While this would not circumvent visual jumps during scene transitions on the individual actor, it would have created a consistent visual timeline for the rest of the scene. In this specific application, this would have provided little advantage, as the scene itself was, intentionally, static and in most cases only one other person was visible at any time, but would have been the most likely approach to take for dynamic environments, such as a busy street scene.

It was also considered to use a 'digital double' approach with a two- or three dimensional model on which the video was mapped and which would algorithmically create the transition between 'neutral' and 'statement' scenes. Current developments in this area, such as the Hyper360 volumetric 3D capturing or the Neon "artificial humans"are promising, but currently not suited for TV use. TVs lack the capability of handling complex models, so any processing and compositing would need to be done on the server side. While this approach is feasible in limited participation test scenarios, it does not yet scale for the potential size of broadcast audiences.

6. CONCLUSIONS AND OUTLOOK

Creating interactive narratives for 360° videos provides challenges specific to that format, as techniques employed in traditional videos, primarily cuts, cannot be employed freely without disrupting the immersive experience.

TV devices have too many restrictions to provide high quality immersive experiences outside a narrow set of constraints. Two technical developments promise a wider range of narrative 360° TV experiences, namely the capability to compose multiple shaped videos on a TV device and cloud-based video with 3D object composition. Both are currently not suited for broadcasters serving large audiences, as there is no sufficient device base for the former and not sufficient scalability for the latter.

With careful planning of the scenarios and the production process, engaging interactive narratives can be created now, provided the current limitations are considered early and the scenario is sufficiently flexible.

7. ACKNOWLEDGMENT

All images include screenshots of rbb productions made specifically for the Hyper360 project.

This work was supported by and received funding from the European Union's Horizon 2020 Framework Programme project Hyper360, under Grant Agreement no. 761934.

8. REFERENCES

[1] https://en.wikipedia.org/wiki/Interactive_television [Apr 2020]

[2] https://en.wikipedia.org/wiki/Dragon's_Lair [Apr 2020]

[3] http://www.quotenmeter.de/n/75991/

[4] http://www.quotenmeter.de/n/61663/

[5] https://www.businessinsider.com/netflix-choose-your-own-adventure-shows-puss-in-boots-and-buddy-thunderstruck-2017-6

[6] https://en.wikipedia.org/wiki/Black_Mirror:_Bandersnatch [Apr 2020]

[7] https://www.hyper360.eu/

[8] Manfredi, A, Patz, N.Tonoli, S., Mihalas, E., Zotta, L. (2018) Hyper360 Requirements and Scenarios, Project Deliverable

[9] L. Bassbouss, S. Steglich and S. Braun, "Towards a high efficient 360° video processing and streaming solution in a multiscreen environment," 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Hong Kong, 2017, pp. 417-422. doi: 10.1109/ICMEW.2017.8026237

[10] Aarseth, E. (2012). A Narrative Theory of Games. In Proceedings of the International Conference on the Foundations of Digital Games - FDG2012. Raleigh, NC, (p. 129-133). doi: 10.1145/2282338.2282365.

[11] Camanho, M.M., Ciarlini, A.E.M., Furtado, A.L., Pozzer, C.T., Feijo, B., 2009. A model for interactive TV Storytelling. In: VIII Brazilian Symposium on Games and Digital Entertainment, Rio de Janeiro, Brazil (SBGames 2009), pp. 197-206, 2009. doi: 10.1109/SBGAMES.2009.31