

Samuel Beckett in Virtual Reality: Exploring narrative using free viewpoint video

Néill O'Dwyer (artist/researcher), Dept. of Drama, Samuel Beckett Centre, Trinity College Dublin (TCD), College Green, Dublin 2, Ireland. Email: odwyernc@tcd.ie. ORCID: 0000-0002-5477-2466.

Nicholas Johnson (director/lecturer), Dept. of Drama, Samuel Beckett Centre, Trinity College Dublin (TCD), College Green, Dublin 2, Ireland.

Enda Bates (sound engineer), Dept. of Electronic & Electrical Engineering (TCD).

Rafael Pagés (computer scientist), Dept. of Computer Science & Statistics (TCD).

Konstantinos Amlianitis (computer scientist), Dept. of Computer Science & Statistics (TCD).

David Monaghan (computer scientist), Dept. of Computer Science & Statistics (TCD).

Aljoša Smolic (principal investigator), Dept. of Computer Science & Statistics (TCD).

Manuscript submitted 25 September 2018.

© ISAST

<1>Abstract

Building on a poster presentation at Siggraph 2018 [1], this article describes an investigation of interactive narrative in virtual reality (VR) through Samuel Beckett's theatrical text *Play*. Actors are captured in a green screen environment using free-viewpoint video (FVV). Built in a game engine, the scene is complete with binaural spatial audio and six degrees of freedom of movement. The project explores how ludic qualities in the original text elicit the conversational and interactive specificities of the digital medium. The work affirms potential for interactive narrative in VR, opens new experiences of the text, and highlights the reorganisation of the author–audience dynamic.

<1>Introduction

What does it mean to be a theatre-maker and an audience in the contemporary digital world? This paper arises from a three-way interdepartmental collaboration on a series of practice-as-research experiments at the intersection of the performing arts and cutting-edge computer-vision and spatial audio techniques. The experiments revolve around a reinterpretation of Samuel Beckett's text entitled *Play* (1963) for virtual reality (VR), and culminated in the realisation of a theatrical VR production, entitled *Virtual Play* [2]. Our goal was to reproduce the play according to the modified conditions of making and watching digital media content in VR.

At the heart of this paper is the question of the performativity of the technology. This refers to the agency of the technology, in terms of its ability to produce a desired result and to reorganise the conditions of production and reception. By harnessing rapidly evolving digital capture techniques and human-computer interfaces, this project shows how they modify the conditions by which knowledge (in this case Beckett's text) is concretised, preserved, translated and handed down to new generations of art publics, who increasingly expect content communicated in ways appropriate to the digital age.

The objectives of the project were three-fold: firstly, to conceive a project that would employ the cutting-edge digital capture technique of free-viewpoint video (FVV) [3]; secondly, to ensure that the production is accessible via VR head-mounted displays (HMDs); finally, to explore storytelling by eliciting the specificities of digital technology – interactivity, dialogue and control.

<1>Rationale

The decision to employ the text of *Play* for the experiment arose out of a multi-year investigation of the text begun by the director [4]. Through a series of interdisciplinary experiments with media artists, computer scientists and sound engineers, this led to a live broadcast *intermedial* version (April 2017) and then to this *virtual* version (October 2017).

Beckett's original text is deeply engaged with the concept of *play*: that is, interactivity and dialogue. His script describes a purgatorial situation wherein three characters, encased in 'three identical grey urns... about 1 yard high' [5], are doomed to repeat the story of their love triangle in perpetuity [6]. They recite the text without direct acknowledgment of each other, in a sequence of articulation prescribed by Beckett. Instead of establishing a dialogic relationship with one another, the urns' inhabitants take cues from a spotlight shone on their faces. Speech is provoked through a visceral reflex to the blinding light, which Beckett suggests is an interrogator. Thus, Beckett invents an interactive game between the actors and the light operator, premised on a simple rule-based system that is similar to a computer programme in its operational binary: light on, speak; light off, stop; loop. The inherent interactive, cybernetic qualities of the performance system make it a resonant text for translation to VR. However, in Beckett's theatrical version, designed for the proscenium paradigm, the audience is always predetermined to sit in the safety of the auditorium and passively observe the game unfolding on the stage. In VR, the interactive, ludic qualities of the medium offer new opportunities for reconsidering the role of the audience as active.

<1>Scenography

The aim of the *mise en scène* was to reinterpret Beckett's scene and characters for VR, using 3D graphics and volumetric video techniques. The original script requires the placement of the actors inside urns [7] at downstage-centre. Working with Beckett's texts in digital media opens a tension between the desire to sensitively and accurately represent his vision and the demands of new media. Previously theorised as a 'spectrum of fidelity' in relation to staging his plays, this dynamic inhabits a scale, balancing *authorial intention* and *audience accessibility*, which witnesses the diminishing of the former, as techno-cultural subjectivities exert more pressure on

the latter [8]. In translating *Play* for VR, conceptual and scenographic adjustments are required for it work in the new medium. However, we are sensitive to the debate around proximity to his original stage directions in contemporary restagings. Thus, we placed an emphasis on challenging contemporary audiences, while maintaining the dramaturgy and scenography consistent with Beckett's original vision. We conceived a scene where the user stands at the centre of a virtual rotunda chamber and is surrounded on three sides by the characters, who face 'undeviatingly [towards the centre] throughout the play' – rather than 'front' as Beckett originally stipulated [9]. We also spaced the urns further apart than he devised (see Fig. 1). However, these decisions are justified by the need for users to experience a measurable sensation of movement in an immersive environment that affords six degrees of freedom (6DoF).

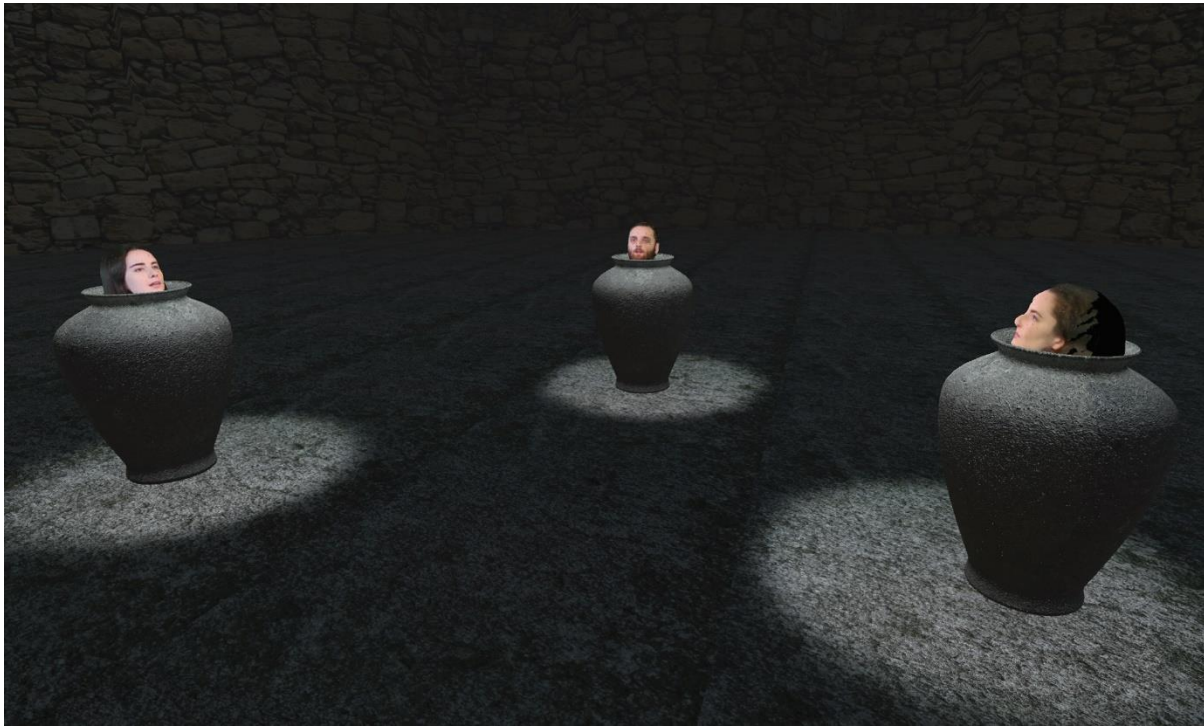


Fig. 1: Virtual rotunda chamber with placement of characters/urns in the triangular relation. © V-SENSE

Another example where affordances and constraints of FVV/VR media led to altered scenographic and directorial decisions is in the characters' faces, which Beckett describes as 'so lost to age and aspect as to seem almost part of the urns' [10]. Film and theatre versions have often responded to this direction by caking the faces in dirt and mud. However, by asking what it means to be 'lost' in digital culture: 'if this is a digital purgatory, rather than a literal post-burial encounter, would it not be the curated, youthful, pure faces reflecting idealised and performed memories, not unlike the Facebook pages of the dead?' [11]. The jitter and flickering in their faces during playback compounds this found-footage aesthetic, an aspect that we chose to keep rather than tidy into a smooth, photorealistic, sterile pallor.

<1>Technical Overview

FVV entails recording actors against a green screen using a multi-camera setup and then stitching the footage together in postproduction. Accurate 3D reconstruction demands maximum

scene coverage and image overlap, so we strategically placed seven DSLRs capturing HD video, in an arc of approximately 150 degrees (see Fig. 2). All captured audio and video are synchronised; the figures are chroma-keyed; the footage is exported as a series of raw images with corresponding silhouettes, and then the camera positions are calculated. These data are used to generate the 3D geometry reconstruction using a shape-from-silhouette (SfS) technique, obtained using the silhouette mask, and a 3D point cloud, acquired through multi-view stereo (MVS). All data are then combined using 3D fusion techniques, resulting in volumetrically complete and accurate 3D models. Input images are used to colour the model using a multi-view texturing technique, producing a photorealistic textured model (see Fig. 3) [12].



Fig. 2: Camera setup in 150-degree arc, with director and actor on set. © V-SENSE



Fig. 3: (from top to bottom): a) reconstructed 3D mesh, before texturing; b) textured model with mesh wireframe; and c) final photorealistic model. © V-SENSE

The immersive nature of VR demands 6DoF spatial audio, wherein the distance and direction of audio sources dynamically change in response to user movements in the scene. This is implemented here using Ambisonics and binaural playback over headphones [13]. Actors' monologues are captured as 24bit monophonic recordings and imported as separate assets into a virtual acoustic environment, powered by the GoogleVR SDK for Unity™ (the game engine). The GVR Room Script is used to approximate the dimensions and materials of the virtual set, and the Source Script is applied to each actor to dynamically alter their direction, distance, and directivity pattern. These ensure that the ratio and timbre of direct and reverberant audio signals change naturally as the user moves further from, or closer to, each character. The overall level of reverberation is also adjusted to ensure sufficient clarity and intelligibility for the dialogue.

The 3D scene, FVV characters [14], and audio are imported into Unity™ and defined as asset bundles. Three nodes are positioned on a 180-degree arc, each containing a custom script that dynamically loads a mesh for each current frame at run-time [15], and one component for playing the (GVR) audio file. The conditions and rules that define user behaviour, interaction and experience are coded into the programme using Unity's C# scripting editor.

<1>Changing the rules: the application of gaming principles to Beckett's text

The computer gaming sector, by employing animation and procedural graphics techniques, has enjoyed a burgeoning of interactive narrative. With the emergence of affordable, high-resolution HMDs, it seems that gaming is the obvious application for VR technology. Film and video, with their dominant language of linear storytelling, struggle to modify strategies or innovate in ways that might open new possibilities for consumable interactive content that still embraces traditional performance methods, like acting and directing for capture technologies. This is precisely the problem that we set out to investigate through our hybrid methodology that encompasses theatre/performance research (art) and creative technology experiments (science).

Notably, as the entertainment industry approaches these technologies, there is an erroneous tendency to conflate 360 video with VR. 360 video is not true VR, because the user's viewpoint is fixed to the camera position stipulated by the moviemaker. There have been some attempts to break away from the pure 360 paradigm by combining it with green screen techniques, e.g. Graham Sack's *Lincoln in the Bardo* (2017), ZDF's *Gladiatoren im Kolosseum*, and Cassandra Herrman and Lauren Mucciolo's excellent, award-winning documentary, *After Solitary* (2017). However, all these are fundamentally constricted by the baked-in viewpoint paradigm and demonstrate little progression on free viewer perspectives explored in *Clouds* (2014) by Jonathan Minard and James George. 'In *Virtual Play* we give the user the sense of natural movement albeit within a defined area as they are allowed to move around the space and interrogate the characters from any point of view' [16], thereby combining the benefits of video-based content creation and 6DoF (naturalistic) user engagement in true virtual environments.

Beckett's text provides a blueprint for how gaming principles can work in the classical performer–audience paradigm, and it impels the exploration of how game-driven narratives can be translated for storytelling in VR. After experimenting with several variations at both the capture and game-building stages, the preference was for a narrative architecture that recognises the interactive specificities of the digital medium, by giving control of the spotlight over to the end user. Several tests confirmed that the most effective method of operation was to attach the spotlight to the user's gaze – wherever the user looks, that space is illuminated (see Fig. 4). When the user looks at the characters, the virtual spotlight shines on them, and they speak. The 'audience' is thus empowered to independently discover the interdependent monologues, simply by looking at the actors and attending to them. This not only has the effect of allowing the audience to determine *who* speaks, but also *when*, and *for how long* [17].



Fig. 4: User point-of-view, looking at character of W1. © V-SENSE

<1>A new way of experiencing *Play*

The new audience-centric approach elicits the interactive specificities of digital technology, and it exposes a consequence that an entirely new way of experiencing the text is possible. In the theatre version, the roving spotlight and ‘toneless’, ‘rapid’ delivery of the text in Beckett’s prescribed order has the effect of casting audiences into a state of hypnosis, wherein the text is perceived as a torrent of words, washing over them; it can be difficult to decipher plot details and individual character perspectives, especially on a first viewing. Beckett seems to tacitly acknowledge this in his final stage direction: ‘Repeat play’ [18]. However, the ability for the *virtual* audience to determine the pattern opens a new experience of the text that is afforded by the reorganised conditions of watching and listening in digital culture. The gameplay is designed so that when one looks away the character stops speaking; when one looks back at them, they pick up exactly where they left off [19]. Users can revisit certain parts of the text and bring renewed focus to that section upon each re-visitation, discover new sections they may have missed during earlier passes, or explore new combinations of textual juxtaposition by editing with their gaze.

This playful and exploratory mode of engagement in *Virtual Play* is only possible using digital database technology. Whereas in analogue technology the recording retains the linear structure of its articulation, in digital databases every sentence, word, syllable and letter become discretised, and therefore equally and immediately accessible. As discrete entities, the bits of information are flattened into a landscape of zero dimension, without horizon or meaning [20]. It is up to the creative team to re-impose meaning through algorithmic design: that is, by rewriting the rules. The power of digital databases has opened great potential for storytelling and narrative, but this potential has yet to be truly innovatively embraced by the performing arts. This is the goal of *Virtual Play*.

In our translation of *Play* for VR, we aim to intensify Beckett's work by allowing users to engage the text on their own terms and absorb it in a playful manner, without changing Beckett's underlying thought. That users can spend as long as they like in the VR environment permits opportunities for deep interpretations of the text that were impossible via traditional observational conditions of theatre. However, users cannot jump around the text as if randomly dragging the playhead of a desktop video player; the user is still bound to listen to each character's individual monologue as scripted by Beckett. Control over the ordering of disclosure is limited to switching between characters, thereby preventing descent into total arbitrariness. While the frenzied quality of Beckett's closed system is clearly diminished in this translation, the effect on the VR user is one of greater ethical responsibility.

An arguable drawback to the open architecture is: because the users are given full control they can decide when to exit the programme, so there is the possibility that one could stop the interrogation process before equally hearing the full testimonies of each character. However, this issue only applies to the first half of each character's text, which is concerned with relating past events. The second half of each monologue is comprised of a series of statements directed at the interrogator (in our case the user). The structure of the later statements is such that they express each character's state of mind [21]. Changing the sequence does not affect their meaning in the context of the overall narrative. Listening to these direct addresses is more akin to engaging in a conversation than listening to a story. Therefore, it challenges the dominant paradigm of telling stories in the past tense; although captured audio-visuals bear an inherent link with the past – with having already happened – engagement with them in VR has the effect of making presence more immediate, thus proposing new opportunities for storytelling.

<1>Conclusion

In the era of mechanical reproducibility, innovations in analogue recording technologies opened new possibilities for storytelling wherein the model of linear narrative – destined for consumption by passive audiences – experienced immeasurable success during the twentieth century. Now, with the rise of increasingly rich, complex, and notably addictive television series, the linear paradigm undoubtedly continues to flourish. While these stories are all distributed and consumed over fibre-optic digital networks giving the illusion of democratisation and choice, they formally employ a unidirectional mode of transmission from active broadcaster to passive viewer. Instead of maintaining user inertia, our conceptualisation of *Virtual Play* aims to help subvert the programme of mass media canalisation, by demonstrating the potential for interactive narratives using cutting-edge technologies of audio-visual capture. This is not an attempt to 'replace' linear narrative – that would be naïve – but to open new possibilities for storytelling that can coexist, complement and supplement the linear incumbents. In *Virtual Play* the inert 'viewer' becomes an energetic 'user', because they have moved from a position of passive reception to one of active exploration.

Despite employing digital video technologies, filmmakers are still broadly locked in an analogue modus operandi, which assumes linear spectating of a singular narrative, under the author's diktat. Opportunities for non-linear narrative were stymied because of shortcomings in playback techniques, but VR offers a structural alternative. Advances in capture, distribution and display

technologies provide new opportunities to alter this imbalance. The difference between Beckett's original linear narrative and our non-linear one can be understood as an example of how evolving technologies and techniques maintain an efficacy that modulates not just the work of art, but also the conditions of its making and reception.

It is important that Beckett's work (and all great literature) continues to be celebrated from one generation to the next. As knowledge is transferred over time, it is interpreted and reactivated in inventive and mutable ways that are wholly answerable to techno-sociological subjectivities and alterations of the conditions of storage, transmission, access, and engagement. Thus, all knowledge undergoes differentiation in its uptake, and *Play* is no exception. We did not set out to make a version of *Play* that contradicts Beckett's stage directions; we aimed to *reactivate* and *represent* his artwork, so that it is accessible to a new generation of art-going publics who expect to engage with culture through the VR medium. In doing so, we elucidate the fluid and transformative qualities of knowledge. We endeavour to strike a balance between celebrating Beckett's landmark exploration of the potential for interactivity in theatre, while also intensifying his work's ideas, by re-interpreting it for VR technologies.

<1>Acknowledgments

Research supported by Science Foundation Ireland Grant #15/RP/2776, Trinity Long Room Hub, and the Estate of Samuel Beckett.

<1>References and Notes

1. N. O'Dwyer et al. "Beckett in VR: Exploring Narrative Using Free Viewpoint Video." In *ACM SIGGRAPH 2018 Posters* (New York: ACM Press) 2018, pp. 1–2.
2. See trailer: https://www.youtube.com/watch?v=0cFYk_1lrG0
3. A. Smolic et al. describe FVV as a volumetric video capture technique that 'offers the same functionality that is known from 3D computer graphics. The user can choose an own viewpoint and viewing direction within a visual scene, meaning interactive free navigation. In contrast to pure computer graphics applications, FVV targets real world scenes as captured by real cameras', "3D video and free viewpoint video-technologies, applications and MPEG standards," in *IEEE International Conference on Multimedia*, 2006, p. 2161.
4. Johnson directed *Play* within *Ethica* (2012–13).
5. Samuel Beckett, *Complete Dramatic Works* (London: Faber&Faber, 2006) p. 307.
6. The final stage direction is 'Repeat play'. Taken in its most logical sense – as a computer programme – this means that upon completion of the text, it loops back to the start and repeats.
7. Although Beckett does not specify a style or material, the urns are known to be funerary in character.

8. N. Johnson, “A Spectrum of Fidelity, an Ethic of Impossibility: Directing Beckett,” in *The Plays of Samuel Beckett* (London: A&C Black) pp. 152–164.

9. See Beckett [5], p. 307.

10. See Beckett [5] p. 307.

11. N. O’Dwyer et al., ‘Virtual Play in Free-Viewpoint Video: Reinterpreting Samuel Beckett for Virtual Reality’, in *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, (2017) p. 265.

12. For a full description of the technical pipeline see: R. Pagés et al., ‘Affordable Content Creation for Free-Viewpoint Video and VR/AR Applications’, in *Journal of Visual Communication and Image Representation* (2018) pp. 192–201.

13. See <developers.google.com/resonance-audio/discover/concepts>, accessed 21 September 2018.

14. One full 3D textured model for every frame of video played back at 30fps equates to c.6,300 models per character, for the 3.5-minute monologues of each actor.

15. This script was necessary, because the file size of the total application was about 6.9Gb, which was crashing the programme; this novel solution is an example of the contribution that aesthetic exigency can make to scientific development.

16. See O’Dwyer et al. [11], p. 263.

17. Beckett opens the sequence to ‘variation’ on the repeat, but in the theatre this is a choice for the director, not the audience.

18. See Beckett [5], p. 317.

19. We considered other formats, including random entry points, having the character begin again, or all monologues continuing at an almost inaudible level when not being watched.

20. Lev Manovich gives a useful account of technical and cultural specificities of databases in *The Language of New Media* (Boston: MIT Press, 2001) pp. 218–236.

21. W1 wishes not to be seen, W2 wishes to be seen, and M craves silence.