Interactive Characters for Virtual Reality Stories

Joan Llobera
Artanim Foundation
Geneva, Switzerland
joan.llobera@artanim.ch

Caecilia Charbonnier
Artanim Foundation
Geneva, Switzerland
caecilia.charbonnier@artanim.ch

ABSTRACT

Virtual Reality (VR) content production is now a flourishing industry. The specifics of VR, as opposed to videogames or movies, allow for a content format where users experience, at the same time, the narrative richness characteristic of movies and theatre plays with interactive engagement. To create such a content format some technical challenges still need to be solved, the main being the need for a new generation of animation engines that can deliver interactive characters appropriate for narrative-focused VR interactive content. We review the main assumptions of this approach and recent progress in interactive character animation techniques that seems promising to realise this goal.

Author Keywords

Interactive Characters; Virtual Reality; Storytelling;

CCS Concepts

•Human-centered computing → Human computer interaction (HCI); Virtual reality;

INTRODUCTION

In the past we have argued [8] that Virtual Reality (VR) opens the possibility to develop a new kind of content format which is, on one hand, experienced as a rich narrative, just as movies or theatre plays are while, on the other hand, it allows interaction with autonomous virtual characters. The experience would feel like having different autonomous characters unfold a consistent story plot, just as a movie or a theatre play does, *within* an interactive VR simulation. Autonomous interactive characters would coordinate to unfold a pre-established script, while still engaging and interacting with users. Autonomous characters would also integrate the actions performed by the users within the unfolding plot, when these actions were relevant for the story.

In such a vision, users would *not* be the protagonist of the story being depicted. Rather, they would only have a part in the story if they performed the relevant actions. However, in all cases, they could still interact and engage with the overall simulation. This approach has the benefit of allowing the merger of a strong plot and an interactive medium, something that has proven challenging in video games. It also allows classical interactive storytelling strategies, like branching structures, but does not need them. VR is rapidly becoming a medium on its own right. What is preventing VR creatives from delivering such experiences? In this paper, we argue that this goal can only be accomplished with the adoption of innovative solutions for interactive character animation, and outline some of our ongoing efforts in that direction.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ACM International Conference on Interactive Media Experiences June 21–23, 2021, Virtual Event, NY, USA

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8389-9/21/06.

DOI: https://doi.org/10.1145/3452918.3465494

VIRTUAL REALITY PSYCHOLOGY

We know that some aspects of VR are fundamentally different from traditional media: people in VR perceive it like a social environment, one in which they adopt a social role when they face a virtual character, and where they have social reactions similar to the real ones [13, 20]. This makes VR very useful for psychological therapies, such as treating fear of heights [4] or for post-traumatic stress disorder [17]. This fundamental difference may be caused by two cognitive phenomena [19]. The first one, called *Place Illusion* is induced by the fact that the visual perspective and the position of the sound sources is updated consistently with your movements. Nowadays, this principle is well understood, and is integrated within the commercial hardware available and the tools to produce VR content. The second, called *Plausibility*, is less well understood, although some empirical work has shown that it is a multi-layered cognitive phenomena [18]. From a storytelling perspective, narrative coherence and interactivity both contribute to the *Plausibility* of the VR experience [8]. It therefore seems desirable to try combining both to get a richer and more plausible VR experience.

A PRODUCTION PERSPECTIVE

The quality and sophistication of VR experiences has radically improved in the last 5 years. The experiences with a broader user base are thought, designed and produced as video games, although omni-directional video and to a limited extent, productions based on volumetric capture, have also increased. Where the first are based on engaging interaction and particularly, often body-centred interaction, the second are focused on narrative development. There are now hundreds of content experiences both from home-VR and location-based VR. There are also more and more VR theater experiences with live actors.

However, we are not aware of a production designed as a theatre play, where different interactions can occur between the virtual characters and the users, within an interactive simulation. Some multi-user experiences do propose short narrative experiences within an interactive virtual simulation (see, for example, Dragons flight Academy[1], Geneva 1850[2] or Aboard The Mayflower[3], depicted in figure 1). There is also considerable interest in engaging directly with virtual characters, and some productions for home consumption have explored this idea.[4] However, these productions remain far from engaging intensively in interaction with virtual characters, as it is routinely done in VR games designed, for instance, in a sports simulator.[5] What we aim at is to have an experience similar to being inside a theatre play, rendered in VR, where we can interact with the virtual characters, but where these characters are automated.

Figure 1. Snapshot of a VR production with virtual characters

A CHALLENGING SET OF TECHNICAL REQUIREMENTS

The main challenge is, in our opinion, a technical one, and one caused by the way animation engines work. The coordination of interactive virtual characters to unfold a story plot is, technically, relatively simple : it is enough to process a story plot like a plan (i.e., a partially-ordered set of events with temporal constraints), and use it as a template for the coordination of goal-motivated artificial agents [3, 10].This can be done by adding the satisfaction of the plan as a goal common to all agents [7]. However, rendering the actions derived from the decisions of goal-motivated actions reveals to be quite more challenging: if we script a story-like plot, it is quite likely that we will consider actions such as opening a door to enter a room, or sitting on a chair. These actions are extremely common in narrative content, such as a movie or a theatre play, but they are uttermost complicated to render using traditional techniques for interactive character animation. Character animation in game engines is typically done using preexisting animations, stored as cinematic trajectories. Therefore, precise interaction with physical objects, like the ones required to open a door, or sit on a chair, is only possible if the set up is completely adjusted in advance, in order that the animation fits with the spatial layout of the door or the chair. This means that the result is not an interactive simulation.

We therefore need an animation engine that can render these kind of actions, but also offers some "wiggle room" to integrate them within an interactive simulation. Such an animation engine would, for example, render an animation like a character opening a door, but doing so with a certain flexibility: it would allow the character to start from different positions and orientations, and to synthesise the animations with slightly different speeds. Such a tool in a production environment would allow the synthesised animation to adjust to temporal and spatial constraints introduced by the actions of another character, or a VR user. Such an animation engine would also be able to render a character sitting on a chair, but in a way that is flexible enough to adapt to slightly different positions and orientations of a chair.

With traditional animation engines it is not possible to introduce real-time adjustments to satisfy with such temporal and spatial constraints, as well as with multiple contacts with physical objects. Additional requirements that such a system would need to satisfy to be effective in a production environment are: first, to synthesise animations consistent with the style of a reference animation, provided by an actor or an animator. Second, to be reasonably fast, as to allow integration within a production pipeline. And third, to work with the existing production tools for virtual reality production, mainly video game engines such as Unreal Engine or Unity3D.

IMPLEMENTATION STRATEGIES AVAILABLE

Physics-based animation

To integrate an interactive character with a virtual environment defined by physical objects the most straight-forward strategy is to use physics-based animation. Peng [14] showed that it was possible to have a physics-based character that imitated the style of a reference animation. Previously, specialised systems had demonstrated the use of physics-based controllers trained with deep reinforcement learning for specific and challenging tasks such as basketball [6], but to our knowledge Peng was the first to introduce an approach that could work for *any* animation. A particularly interesting aspect of this method is that it is simple: since it was based on pure reinforcement-based learning, with no underlying model, there were no particular assumptions on the topology of the joints, nor on the kind of movement, nor on the dynamics of the movement imitated. The authors also showed that introducing a random parameter position at the training phase, with a positive reward for reaching it either with the body or with a limb, or even with a ball. The same authors also showed [16] it could work using reference animations extracted from a simple monocular video. This means it could avoid the need to use high-end motion capture systems.

A limitation of the previous method is that each movement has to be synthesised with a policy trained for a particular movement. Therefore, combining animations can be done using a mixture of experts approach, but these methods rapidly become difficult to scale, because each kind of behaviour requires a separate controller. Distillation techniques have been

[1] https://dreamscapeimmersive.com/adventures/details/dragons01
[2] https://artanim.ch/project/geneva-1850/
[3] https://artanim.ch/project/mayflower-vr/
[4] To our knowledge the best example publicly available is Wolf In The Walls https://www.oculus.com/experiences/rift/2272579216119318/
[5] see, for example, Eleven Table Tennis https://www.oculus.com/experiences/rift/989106554552337/

used to train a more general network that learns to imitate a large set of specialised controllers [11, 12], borrowing from cognitive sciences the notion of *motor primitive*. An alternative strategy proposed has been to consider each policy made of more primitive policies, and then combine them, much like a gaussian mixture model is defined by the multiplication of simpler gaussian distributions [15].

A different strategy for physics-based animation consists in combining a cinematic controller with a physics-based controller that integrates the movements generated by the cinematic controller with the constraints of the physical simulation. [1] showed motion matching (introduced in [2]) and a physics controller can be combined. This gives the flexibility of animation methods driven by cinematic data with the benefits of physics-based animation. [23] further extended this approach proposing a universal physics-based animation layer, trained with a very large database of humanoid animations and different cinematic controllers. The fundamental difference here is that the cinematic controller is still used in the inference step: the cinematic controller generates a reference pose, integrating also high-level input, and the policy merely imitates it. This suggests a strategy of divide and conquer: if it is possible to develop a universal physics controller to integrate an interactive character within a physics scene, then the generation of the behaviour of the interactive character can be achieved with *any* cinematic controller.

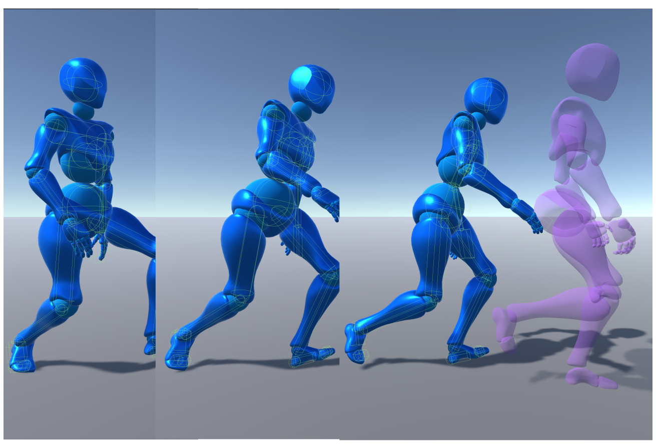

Figure 2. A rigged character, in blue, driven by a rag doll made of linked rigid bodies, shown as green silhouettes. The rag doll is controlled by a physics controller. The physics controller uses a reference cinematic animation (here depicted in purple) to generate movements that are consistent with the style of a reference animation.

Cinematic controllers

Motion Matching [2] was introduced as a new way to combine motion data that provided smooth transitions. It has already been implemented in commercial video games, and implementations are available off the shelf. However, motion matching works mainly for navigation tasks, and has difficulty generalising to different tasks. An approach that has shown fruitful to develop innovative controllers is to pre-process the motion data to extract motion descriptors, and then use these to guide a neural controller. For example, [5] showed that for cyclic movements the extraction of the phase could be used to guide a novel kind of neural network to synthesise appropriate movements in a navigation task, following reference animations. This solved some of the challenges found when trying to train neural controllers with recursive neural networks. Starke [21] demonstrated how to train a generative network to synthesise cinematic animations that require complex interactions with the environment, such as sitting on a chair, or going through a door. This kind of method use supervised learning to train a generative animation controller, using the trajectories of different limbs, as well as a multi-dimensional representation of the space surrounding the agent. This allows implementing abstract controllers to synthesise behaviour. A significant development has been the notion of *local motion phases*. Instead of considering the phase of a cyclic movement for the entire body, [22] showed a phase can be defined for any limb, between two contacts. As a result, he demonstrates a generic approach to implement cinematic controllers that require complex movements, like a basketball scenario, which previously could only be addressed with specialised controllers. The authors further demonstrate that local motion phases improve the neural controller used to synthesise complex interactions: the way characters move through doors, sit on chairs, and interact with other advanced controllers becomes smoother, and more effective. The main limitation remaining is that local motion phases can only be applied for movements that involve physical contact. Therefore, when we consider spatial constraints that do not require contact, a solution is still needed.

DISCUSSION

Our goal is to develop stories in VR based on interactive characters, and unfold a plot by coordinating their behaviour. This requires an innovative approach to character interaction. In our opinion the simplest way to do this is to combine a physics-based animation system with a data-driven cinematic controller.

Why not use a purely physics-based animation strategy? Physics-based animation is generally based on learning a policy using deep reinforcement learning. This means that the reward is provided by the physical simulation, and they do not need labelled data. Physics-based animation based on motor primitives or compositional motor policies effectively solve the problem of scaling physics-based animation strategies when a scenario requires a multiplicity of specialised controllers. However, this can come at the cost of not preserving the original style of the motions provided, which can be a major obstacle to the delivery of the intention portrayed by the actor or the animator. The use of a purely physics-based controllers also seems difficult to combine with controlling the temporal dynamic of the animation, something much easier to do with cinematic-based controllers.

A more practical aspect of using physics-based animation systems for interactive VR characters is that generally these solutions are are not implemented in tools available for game production. Some of them use the physics engine Bullet, and some are developed in proprietary platforms, without providing a public implementation. Moreover, physics-based animation almost universally uses rag dolls, and not rigged characters, which are mandatory for VR and video game productions. Although these are not fundamental limitations of these techniques, these practical aspects add further challenges to their adoption in VR productions.

Given the tools available for interactive character animation, we believe the most sensible strategy is to combine a system similar to the one proposed in [21] and [22], together with a physics layer that takes care of the integration with the physics of the simulation, and integrates collisions and physical contacts in the animation. The limitation on having local phases defined based on physical contacts is still a challenging one, but for a significant amount of the animations needed, such a system could work. And further improvements can be developed to address those. This, combined with a physics-based controller similar, in an approach similar to the one described in [1], would strongly simplify the development of autonomous virtual characters.

Additional requirements would be to have a system that works with rigged characters, and is easy to integrate in a VR production environment. Towards that direction we have recently developed a re-implementation of the system described in [1], but available in Unity3D, a game engine that provides many tools that simplify the development of VR experiences (see figure 2).[6] Our implementation works with rigged characters, and we have tested it works with the default animation system available in this engine, Mecanim, as well as with an implementation of motion matching that can be licensed at an affordable price.[7] An introductory course will be available soon (see [9]). The following logical step seem to be to combine this data-driven physics controller with innovative cinematic controllers, and explore to what extent, when combined with the solution for plot coordination previously described, such a technical architecture can be used to unfold a story in VR based on the coordination of users and automated interactive characters.

An open animation challenge beyond the scope of this solution is to develop a solution to render the kind of subtle interaction that people spontaneously expect when facing a character of their size, in a shared virtual space. This implies exploring methods to render spontaneous interpersonal synchronisation mechanisms that we establish, both unconsciously and willingly, in our everyday interactions to denote rapport, complicity, hostility, and a variety of subtle message that reflect on the intent we attribute to others, and that others attribute to ourselves. We hope our current efforts help spark a wider interest in this research and engineering path.

REFERENCES

[1] Kevin Bergamin, Simon Clavet, Daniel Holden, and James Richard Forbes. 2019. DReCon: data-driven responsive control of physics-based characters. *ACM Transactions On Graphics (TOG)* 38, 6 (2019), 1–11. DOI:http://dx.doi.org/10.1145/3355089.3356536

[2] Simon Clavet. 2015. Motion Matching - The Road to NextGen Animation. *Proc. of GDC* (2015). https://www.gdcvault.com/play/1023280/Motion-Matching-and-The-Road

[3] Klaus Dorer. 2004. Extended behavior networks for behavior selection in dynamic and continuous domains. In *Proceedings of the ECAI workshop Agents in dynamic domains, Valencia, Spain*. Citeseer.

[4] Daniel Freeman, Polly Haselton, Jason Freeman, Bernhard Spanlang, Sameer Kishore, Emily Albery, Megan Denne, Poppy Brown, Mel Slater, and Alecia Nickless. 2018. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry* 5, 8 (2018), 625–632. DOI: http://dx.doi.org/10.1016/S2215-0366(18)30226-8

[5] Daniel Holden, Taku Komura, and Jun Saito. 2017. Phase-functioned neural networks for character control. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–13. DOI:http://dx.doi.org/10.1145/3072959.3073663

[6] Jessica Hodgins Libin Liu. August 2018. Learning Basketball Dribbling Skills Using Trajectory Optimization and Deep Reinforcement Learning. *ACM Transactions on Graphics* 37, 4 (August 2018). DOI: http://dx.doi.org/10.1145/3197517.3201315

[7] Joan Llobera. 2012. *Stories within Immersive Virtual Environments*. Universitat de Barcelona.

[8] Joan Llobera, Kristopher J Blom, and Mel Slater. 2013. Telling stories within immersive virtual environments. *Leonardo* 46, 5 (2013), 471–476. DOI: http://dx.doi.org/10.1162/LEON_a_00643

[9] Joan Llobera, Joe Booth, and Caecilia Charbonnier. 2021. New Techniques in Interactive Character Animation. In *SIGGRAPH '21 Courses, August 09-13, Virtual Event, USA,Special Interest Group on Computer Graphics and Interactive Techniques*. DOI: http://dx.doi.org/10.1145/3450508.3464604

[10] Joan Llobera and Ronan Boulic. 2017. A tool to design interactive characters based on embodied cognition. *IEEE Transactions on Games* 11, 4 (2017), 311–319. DOI:http://dx.doi.org/10.1109/TCIAIG.2017.2755699

[11] Josh Merel, Leonard Hasenclever, Alexandre Galashov, Arun Ahuja, Vu Pham, Greg Wayne, Yee Whye Teh, and Nicolas Heess. 2019. Neural Probabilistic Motor Primitives for Humanoid Control. In *International Conference on Learning Representations*.

[12] Josh Merel, Saran Tunyasuvunakool, Arun Ahuja, Yuval Tassa, Leonard Hasenclever, Vu Pham, Tom Erez, Greg Wayne, and Nicolas Heess. 2020. Catch & Carry: reusable neural controllers for vision-guided whole-body tasks. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 39–1. DOI: http://dx.doi.org/10.1145/3386569.3392474

[13] Xueni Pan, Marco Gillies, and Mel Slater. 2015. Virtual character personality influences participant attitudes and behavior–an interview with a virtual human character

[6] see https://joanllobera.github.io/marathon-envs/

[7] https://assetstore.unity.com/packages/tools/animation/motion-matching-for-unity-145624

about her social anxiety. *Frontiers in Robotics and AI* 2 (2015), 1. DOI: http://dx.doi.org/10.3389/frobt.2015.00001

[14] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. 2018. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–14. DOI: http://dx.doi.org/10.1145/3197517.3201311

[15] Xue Bin Peng, Michael Chang, Grace Zhang, Pieter Abbeel, and Sergey Levine. 2019. MCP: Learning Composable Hierarchical Control with Multiplicative Compositional Policies. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*.

[16] Xue Bin Peng, Angjoo Kanazawa, Jitendra Malik, Pieter Abbeel, and Sergey Levine. 2018. Sfv: Reinforcement learning of physical skills from videos. *ACM Transactions On Graphics (TOG)* 37, 6 (2018), 1–14. DOI:http://dx.doi.org/10.1145/3272127.3275014

[17] Albert Rizzo, Greg Reger, Karen Perlman, Barbara Rothbaum, JoAnn Difede, Rob McLay, Ken Graap, Greg Gahm, Scott Johnston, Rob Deal, and others. 2011. Virtual reality posttraumatic stress disorder (PTSD) exposure therapy results with active duty OIF/OEF service members. *International Journal on Disability and Human Development* 10, 4 (2011), 301–308. DOI: http://dx.doi.org/10.1515/IJDHD.2011.060

[18] Richard Skarbez, Solene Neyret, Frederick P Brooks, Mel Slater, and Mary C Whitton. 2017. A psychophysical experiment regarding components of the plausibility illusion. *IEEE transactions on visualization and computer graphics* 23, 4 (2017), 1369–1378. DOI: http://dx.doi.org/10.1109/TVCG.2017.2657158

[19] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557. DOI: http://dx.doi.org/10.1098/rstb.2009.0138

[20] Mel Slater, Aitor Rovira, Richard Southern, David Swapp, Jian J Zhang, Claire Campbell, and Mark Levine. 2013. Bystander responses to a violent incident in an immersive virtual environment. *PloS one* 8, 1 (2013), e52766. DOI: http://dx.doi.org/10.1371/journal.pone.0052766

[21] Sebastian Starke, He Zhang, Taku Komura, and Jun Saito. 2019. Neural state machine for character-scene interactions. *ACM Trans. Graph.* 38, 6 (2019), 209–1. DOI:http://dx.doi.org/10.1145/3355089.3356505

[22] Sebastian Starke, Yiwei Zhao, Taku Komura, and Kazi Zaman. 2020. Local motion phases for learning multi-contact character movements. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 54–1. DOI: http://dx.doi.org/10.1145/3386569.3392450

[23] Tingwu Wang, Yunrong Guo, Maria Shugrina, and Sanja Fidler. 2020. UniCon: Universal Neural Controller For Physics-based Character Motion. (2020). https://nv-tlabs.github.io/unicon/