# Towards computationally efficient VR agents

Magdalena Igras-Cybulska
migras@agh.edu.pl
Institute of Electronics
AGH University of Science and Technology
Kraków, Poland

Artur Cybulski
kontakt@upgrade.com.pl
Upgrade Artur Cybulski
Kraków, Poland

Radosław Sterna
radoslaw.sterna@gmail.com
Emotion and Perception Lab
Institute of Psychology
Jagiellonian University in Kraków
Poland

Marta Bartoszek
marta.bartoszek2525@gmail.com
Institute of Psychology
Jagiellonian University in Kraków
Poland

Agnieszka Siry
agasiry@gmail.com
Institute of Psychology
Jagiellonian University in Kraków
Poland

## ABSTRACT

Generating multiple bots which look and move authentically and at the same time differently from each other is a difficult task. While working on it, researchers need to deal with problems of low time-efficiency and complexity of the process as well as trade-off between realism of a character and their generation time, while hardware performance should be also kept in mind. We present our solution which consists of pipelines for generating bots appearance and behaviour as well as optimization steps.

## KEYWORDS

virtual agents, render pipeline, VR, agent performance, Unity, NPC

**ACM Reference Format:**
Magdalena Igras-Cybulska, Artur Cybulski, Radosław Sterna, Marta Bartoszek, and Agnieszka Siry. 2021. Towards computationally efficient VR agents. In *MIG'21: Motion, Interaction and Games.* ACM, New York, NY, USA, 3 pages. https://doi.org/XXX

## 1 INTRODUCTION

Creating a psychological experiment with a virtual character is a challenging task itself and the difficulty level only increases with the characters number. Generated multiple bots, including crowds, need to seem authentic and at the same time diversely looking. To make them look natural, the characters often need to have different shaders, textures and their parameters to provide a diversity in their appearances. They need to move in a realistic and situationally adjusted way, including face motions and eye control. Moreover, their movements need to be non-identical to prevent a sense of artificiality. As Pan and Hamilton [8] note, until now the researchers struggled with character design due to its low time-efficiency and complexity while achieving a minor number of characters. Moreover, there is a significant trade-off between realism of the characters and the time necessary for their creation. As an outcome, the researchers often prioritize the time-efficiency which results in lower realism achieved [14]. In consequence, a trend of research emerged aiming at exploring the possibilities of automatization of virtual character creation [5]. It might reduce the time needed for character construction [1] and increase accessibility of virtual character design process [3]. Another problem is that when the stimuli number increases the probability of a glitch occurrence grows, which might lead to a simulator sickness [2]. While working with a big number of characters, the researchers thus need to bear in mind hardware performance. Yet, currently, there is a scarcity of such comprehensive solutions [9]. Some of the VR researchers attempted creating pipelines of virtual character creation. For example, Grigorev et al. [4] proposed an approach called neural dressing which allows for imposing clothing and hair on the body mesh. Based on it, they created a generative model of full-body avatars, which combines polygonal body mesh modeling with neural rendering and enables obtaining avatars from one or several pictures of the person. Piao et al. [9], aiming at recreating a large-scale crowd of virtual avatars via the internet, used Web3D technology and combined model parameterization technique (which parameterizes the bone shapes, texture and animations) based on shape space with a multi-level clone instantiation.

VR applications for public speaking training are an example case where realism of agents behavior can directly affect the acceptance of the method and the effectiveness of training [7]. Among main design challenges that are posed to the customization of virtual audience, identified by [13], there are realism of the visual representation, randomization of audiences' looks and behaviors for the sake of realism. Social anxiety induced by VR environment was impacted also by the size of virtual audience [6]. Although the topic exists in the literature, we think that there is still a need for developing novel automatization solutions. As a result of encountering these problems ourselves, we would like to present the pipeline we developed, which helps to automate the process and avoid the aforementioned difficulties. It enables generating multiple virtual characters which exhibit diverse appearances and behaviors, without overstraining the hardware.


*MIG'21, November 10–12, 2021, Lausanne, Switzerland*

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXX

## 2 PIPELINE OF AGENTS PRODUCTION

**Agents appearance.** Firstly, all of the bot variants were created in *Character Creator 3* (CC3) with use of photogrammetric scans with texture resolutions mainly 4096x4096 pixels. Then, with use of one of the agents moved to the *iClone 7* and also with use of iPhone *Live Face* app on iPhone 11 Pro - the facial motion capture recordings were made. The sessions were additionally aided with concurrent use of realtime audio recording using the Zoom H6 as a microphone for both improving the facial motion capture and recording the speech to be played as a sentence said by the bot. For the facial motion capture part - one of the most advanced at the time of the development markerless methods was used - Apple's ARKit blendshapes profile with *iClone's 7.9 Live Face* motion capture combined with Live Smooth adjustments resulted in quite compelling facial motion capture *iMotionPlus* animation recordings embedded separately alongside *CC3* bot variant previously converted to *iMotionPlus*-compatible CC+ standard - into .fbx file exported with Unity export settings preset - exported from within *CC3* for widest variety of the exported texture types. The .fbx files holding the bot variant model and the animation alongside the exported non-embedded textures were then imported into the Unity project dedicated for research environment creation. Then it was adjusted with various improvements of which the most important were the shaders used for the highest visual quality with maintaining the best possible performance (specifically for most fluent rendering that would match the maximum Vive Pro-allowed framerates of 90fps) needed for later to be able tu run the biosignals processing on the same PC at the time of VR rendering during the research session.

In order to obtain significant differences between three visual quality categories of the bot variants (the levels of perceived realism were investigated in [11]), the following shader and texture settings were used: 1) highest realism: shaders: *Apollo URP Translucent Opaque*, *Lux URP Human Hair Blend*, *URP Complex Lit*, *Eye Advanced URP*; textures: mostly 4096x4096, an example is presented in Fig. 2; 2) medium: shaders: as above, but with no smoothness and no translucency; textures: fixed resolution 256x256 each, 3) low: shaders: *URP unlit*; textures: fixed resolution 64x64 each.

**Agents behavior.** The behavior of the bot variants consisted of animations for specific body parts with the rest of the body masked for such animation. This way the Unity animator for each variant had various layers with complementarily masked animations - separate layer for facial mocap and separate ones for animations of the rest of the body. The eyes and head movements were managed programmatically by *Realistic Eye Movements* Unity asset.

**Optimization.** For achieving the high performance while keeping high quality the custom-adjusted Universal Render Pipeline (URP) was used. For necessary, but also possibly lightweight visual postprocessing the Sharpness of the *Beautify 2* Unity Asset was applied - significantly increasing the visible detail. Optimization methods included sharing all repetitive textures and GPU-instancing-enabled materials between variants, for displaying the variants the pooling was used (enabling, and then disabling the Unity scene GameObjects holding randomly displayed variants). In order to reduce the stuttering the custom interpolation method was used [10].

## 3 EVALUATION

For evaluation purposes we generated 10 different variants of appearance of agents in the highest realism settings. The tests of framerate as a function of number of bots visible in the scene (Fig. 1) were performed in the editor and not in the build, so the expected final performance will be higher. We used test build: GeForce RTX 2080Ti, Ryzen 9 3900X, RAM 64GB, NVMe SSD in setup for HTC Vive Pro headset.

## 4 DISCUSSION AND CONCLUSIONS

The biggest challenge that we met was obtaining the balance between the highest possible visual, behavioral and interactional quality while keeping the lowest possible impact on the hardware performance. Additionally, we intended to leave the biggest possible reserve to the biometric signals acquisition and processing (exemplary setup was described in [12]). Picking proper Unity render pipeline, appropriate set of shaders and their settings was crucial. If the visual quality would be the only determinant then Unity's HDRP would be the logical choice. But because of the above mentioned performance buffer priority, the customized URP with *Beautify 2* sharpness and art filmic ACES tonemap operator postprocessing were optimal. While both *CC3* and *iClone 7* are the tools providing high quality results in relatively short time, the output imported into Unity requires many adjustments. Realistic eyes and head look and non-hard-coded behavior required to replace the initial eye textures and shader to URP *Eye Advanced*. Eyes and heads were put under control of the *Realistic Eye Movements* with masking out head and eyes animations from all animator layers. One of the bigger challenges was also to obtain the satisfactory hair. The used haircuts assets differed in their properties and thus each different haircut required manual adjustments of different shaders to get proper alpha channel management.

The next optimization steps will involve Streaming Virtual Texturing combined with HDRP using i.e. subsurface scattering, GPU instancing for crowd animations and the "impostors" technique. The described methods are planned to be applied for generating crowds of virtual agents as a virtual audience for VR trainer of public speaking. We will apply emotional expressions captured with actors and perform further perceptual tests with control of immersion and co-presence levels.

The solution presented in this work consists of a pipeline of generating multiple virtual characters. Our work can serve as an inspiration for other researchers on how to setup and configure existing tools in order to obtain a similar effect and we hope that our work will facilitate future studies on effective virtual character creation.

## 5 ACKNOWLEDGMENTS

The paper was supported partly by program Lider nr 0230/L-11/2019 by the National Center for Research and Development and partly by Miniatura grant nr 2020/04/X/ST6/00626 by the National Science Centre, Poland. Work of R. Sterna was financed from the Diamond Grant programme nr DI2018 015848 granted by the Ministry of Science and Higher Education of Poland.

**Figure 2: Example agent**

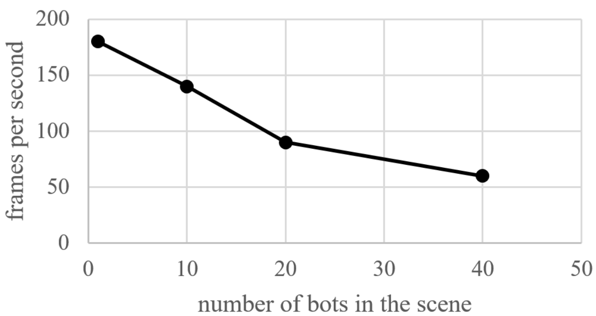


**Figure 1: Performance for multiple bots of the highest quality in the scene**

## REFERENCES

[1] Jascha Achenbach, Thomas Waltemate, Marc Erich Latoschik, and Mario Botsch. 2017. Fast generation of realistic virtual humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 1–10.

[2] Natalia Dużmańska, Paweł Strojny, and Agnieszka Strojny. 2018. Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Frontiers in psychology* 9 (2018), 2132.

[3] Andrew Feng, Evan Suma Rosenberg, and Ari Shapiro. 2017. Just-in-time, viable, 3-D avatars from scans. *Computer Animation and Virtual Worlds* 28, 3-4 (2017), e1769.

[4] Artur Grigorev, Karim Iskakov, Anastasia Ianina, Renat Bashirov, Ilya Zakharkin, Alexander Vakhitov, and Victor Lempitsky. 2021. StylePeople: A Generative Model of Fullbody Human Avatars. *arXiv preprint arXiv:2104.08363* (2021).

[5] Jai Mashalkar and Parag Chaudhuri. 2017. Creating Personalized Avatars. In *Digital Hampi: Preserving Indian Cultural Heritage*. Springer, 283–298.

[6] Fariba Mostajeran, Melik Berk Balci, Frank Steinicke, Simone Kühn, and Jürgen Gallinat. 2020. The effects of virtual audience size on social anxiety during public speaking. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 303–312.

[7] Fabrizio Palmas, Jakub Cichor, David A. Plecher, and Gudrun Klinker. 2019. Acceptance and Effectiveness of a Virtual Reality Public Speaking Training. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 363–371. https://doi.org/10.1109/ISMAR.2019.00034

[8] Xueni Pan and Antonia F de C Hamilton. 2018. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology* 109, 3 (2018), 395–417.

[9] Xue Piao, Yang Li, Kang Xie, Hantao Zhao, and Jinyuan Jia. 2020. Towards Web3D-based Lightweight Crowd Evacuation Simulation. In *The 25th International Conference on 3D Web Technology*. 1–9.

[10] Scott Sewell. [n.d.]. *Timesteps and Achieving Smooth Motion in Unity*. Retrieved May 7, 2021 from https://www.kinematicsoup.com/news/2016/8/9/rrypp5tkubynjwxhxjzd42s3o034o8

[11] Radosław Sterna, Artur Cybulski, Magdalena Igras-Cybulska, Joanna Pilarczyk, and Michał Kuniecki. 2021. Pretest or not to pretest? A preliminary version of a tool for the virtual character standardization. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 123–126. https://doi.org/10.1109/VRW52623.2021.00030

[12] Radosław Sterna, Artur Cybulski, Magdalena Igras-Cybulska, Joanna Pilarczyk, Agnieszka Siry, and Michał Kuniecki. 2021. Psychophysiology, eye-tracking and VR: exemplary study design. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 639–640. https://doi.org/10.1109/VRW52623.2021.00202

[13] Ana-Despina Tudor, Sandra Poeschl, and Nicola Doering. 2013. Virtual audience customization for public speaking training procedures. In *2013 IEEE Virtual Reality (VR)*. 61–62. https://doi.org/10.1109/VR.2013.6549363

[14] Egor Zakharov, Aleksei Ivakhnenko, Aliaksandra Shysheya, and Victor Lempitsky. 2020. Fast Bi-layer Neural Synthesis of One-Shot Realistic Head Avatars. In *European Conference on Computer Vision*. Springer, 524–540.