Humans in (Digital) Space

Representing Humans in Virtual Environments

Mark Lycett*
Department of Digital Innovation and Management, Royal Holloway, University of London, Egham TW20 0EX, Surrey, United Kingdom
mark.lycett@rhul.ac.uk

Alex Reppel
Department of Marketing, Royal Holloway, University of London, Egham TW20 0EX, Surrey, United Kingdom
alexander.reppel@rhul.ac.uk

ABSTRACT

Technology continues to pervade social and organizational life (e.g., immersive, and artificial intelligence) and our environments become increasingly virtual. In this context we examine the challenges of creating believable virtual human experiences— photo-realistic digital imitations of ourselves that can act as proxies capable of navigating complex virtual environments while demonstrating autonomous behavior. We first develop a framework for discussion, then use that to explore the state-of-the-art in the context of human-like experience, autonomous behavior, and expansive environments. Last, we consider the key research challenges that emerge from review as a call to action.

CCS CONCEPTS

• **Human-centered computing** → Human computer interaction (HCI); Interaction design; Collaborative and social computing.

KEYWORDS

Believable Virtual Humans, Digital Humans, Metaverse

ACM Reference Format:
Mark Lycett and Alex Reppel. 2022. Humans in (Digital) Space: Representing Humans in Virtual Environments. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI 2022), June 06–10, 2022, Frascati, Rome, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3531073.3531172

1 INTRODUCTION

Digital replicas that look and act like real humans occupy the imagination of creators, marketers, and technologists alike. Following the literature to-date, we scope our definition of these replicas as complex photo-realistic three-dimensional human models alongside the associated logic capable of delivering believable real-time behavioral responses and emotional interactions across multiple and diverse audiences. From one perspective, these Believable Virtual Humans (BVHs) represent the ultimate advanced visual interface, where they proxy for all human senses in digital interactions supported via artificial intelligence (AI) and immersive technologies (for movement, gesture, touch, etc.).

Some suggest that putting 'a face on technology' makes it more accessible—allowing, for example, the inclusion of nonverbal cues [28]—and those virtual stand-ins will make life both easier and more convenient in relation to social interaction, work, etc. Such optimism stands in contrast to the many fictitious stories that often describe unintended consequences of creating artificial humans. (Examples range from anthropomorphic physical creatures, such as malfunctioning robots in Michael Crichton's Westworld to purely virtual entities, such as HAL 9000 in Arthur C. Clarke's Space Odyssey series.) Moreover, whenever marketers and technologists proclaim that an innovation is on the horizon, it is sensible to explore the gap between hype and reality. That is, to explore the current limits on changes to the way people interact; whether such interfaces/technology will be easier to use; enable experiences that may otherwise not be possible; and whether we will value such experiences and the like.

To do this, we develop an outline framework with which to evaluate the potential of BVHs and categorize the challenges associated with creating and living with them. At this stage, our intention is not to be all encompassing but, rather, to present what we see as the key issues, the potential research opportunities moving forward, and to stimulate lively debate in the context of the conference.

For this, the paper is structured as follows. Section 2 outlines the concept of Believable Virtual Humans (BVH). Section 3 discusses the current state-of-the-art related to human-like appearance, autonomous behavior, and expansive environments. Section 4 presents an outline framework and puts forward a set of propositions. In Section 5, we apply this framework to explore challenges and key research opportunities of the idealized vision of BVHs in the context of the state-of-the-art. Section 6 concludes the work.

2 BELIEVABLE VIRTUAL HUMAN EXPERIENCES

Interest in life-like digital imitations—often referred to as virtual humans [7]—is increasing [1–4, 10, 27], with the most obvious innovation being the creation of life-like photo-realistic CGI [computer-generated imagery] characters with realistic motion controls. While a photo-realistic appearance is an important first step, the emerging commercial goal (fueled by what is now referred to as the 'metaverse') is the creation of:

> "Virtual Worlds [...] where Virtual Humans will co-operate, negotiate, make friends, communicate, group

*Corresponding author.

AVI 2022, June 06–10, 2022, Frascati, Rome, Italy

ACM ISBN 978-1-4503-9719-3/22/06...$15.00
https://doi.org/10.1145/3531073.3531172

> and break up, depending on their likes, moods, emotions, goals, fears, etc. [...] Behavior should emerge as a result of a multi-agent system sharing a common environment [...] Virtual Humans have their own motivations and needs, are able to sense and explore their environment, and their action selection mechanism determines suitable actions to take at any time" [17, p. 2].

This vision goes beyond current implementations of virtual humans or even the most life-like CGI characters used in films. Instead, future virtual humans are described as being capable of navigating complex environments while demonstrating autonomous behavior. Even though many of the necessary technologies already exist, the realization of this ambitious vision is still some way off.

3 STATE-OF-THE-ART TECHNOLOGIES

Instead, the current generation of virtual humans is commonly understood as having a photo-realistic human-like appearance and the potential to fulfil tasks currently done by real humans [24]. Consequently, our survey of the current state-of-the-art includes technologies to ensure human-like appearance and autonomous behavior, which we extend by also exploring the environments in which virtual human interactions take place, culminating in a framework to better understand the potential and challenges associated with BVH experiences

3.1 Human-like appearance

Virtual humans can be described as photo-realistic three- dimensional human-like models that are projected into physical environments (often referred to as holograms) or that represent either users, celebrities, or roles typically associated with humans in the physical world. Consequently, a photo-realistic human-like appearance that is both distinct and adaptable is the first essential component to ensure believable virtual human experiences.

A major feature of the current generation of virtual humans is their (almost) photo-realistic human-like appearance. This is the domain of digital human modelling [6] and the same technologies that help create CGI characters in films are used to create photo-realistic three-dimensional virtual humans. Broadly speaking, digital human modelling involves three main steps: (a) Scanning, (b) modelling, and (c) rendering (animation) [24]—for more detail see [7]. At present, realism is achieved using technologies such as volumetric capture, motion capture and/or in combination with complex animation, which are expensive and often time-consuming. The outcome of the digital human modelling is a digital human or abstract model stored in a computer-readable file (asset) that defines the photo-realistic human appearance of a character shown on a screen or projected into physical space. Digital humans form a super-set of virtual humans and other digital representations: A model can represent digital avatars, holograms, digital doubles, or virtual humans. Digital avatars are representations of real people participating in virtual worlds (games, virtual communities, etc.) and are generally controlled by the player. A realistic almost life-like representation of a real person is sometimes referred to as an interactive digital human [24], a special case of a digital avatar. (An important difference between the two is the way interactive digital humans are controlled, typically by voice translating actions in real-time into movements for the digital human.) Other entities represented by a digital human are holograms that project a model into a physical environment, digital doubles, or replicas/imitation of well-known individuals (celebrities), and virtual humans that represent specific roles in virtual environments.

Where characters can be generic in nature (e.g., customer service), frameworks are starting to appear that simplify their development–Unreal's MetaHuman Creator is an early example. Further, progress is required for capture technologies to move beyond specialist studios [22], and for character animation to move from being scripted to autonomous in nature–i.e., characters respond dynamically to a situation/context. In addition, work is required on autonomous behavior [23].

3.2 Autonomous behavior

Another feature of current virtual humans is their aim to fulfil tasks currently done by real humans. From a traditional computational perspective, the demand here might be seen as a cognitive architecture to enable the acquisition and (appropriate) use of knowledge [14]. Autonomous behavior that supports (visual) adaptability and interactions amongst entities and their environments constitutes the second component of our framework.

A cognitive infrastructure does not (necessarily) suggest an ability to 'think,' but rather the ability to respond (appropriately) to external stimuli. In other words, it defines a virtual human's abilities as driven by artificial intelligence (AI). Although a universally accepted definition of AI appears not to exist [12], it is sometimes colloquially expressed as simulating human intelligence [29]. Such an understanding can be misunderstood as implying some form of human-like capacity for independent thought or even self- awareness. While some predict a future of general AI where self-aware digital entities gain the capacity to think independently [16], current AI is better understood as an attempt to automate at scale, that is to translate learning from data into autonomous behavior.

This has implications for virtual humans. First, it makes it possible to animate photo-realistic digital humans in real-time (the expressive ability of BVHs). Second, it allows virtual humans to gain knowledge of the past (their adaptive ability). Third, it makes it possible for virtual humans to engage in conversations and (inter)actions (their responsive/autonomous ability).

Automation at scale is the domain of machine-learning (ML), an area where significant progress of late has been achieved via deep learning (DL). Based on neural networks, this approach has shown itself effective in several areas—computer vision/pattern recognition, natural language processing, text classification tasks, and dialogue management [15, 18–20]. Indeed, the latter is the technology at the heart of state-of-the-art chatbots, which may be seen as a precursor to virtual humans in conversational terms—its novelty rooted in the maintenance of dialogue and ability to recognize human variance [8]. Further, DL is also at the heart of facial simulation (expressiveness) and use of neuroscientific models of cognition to computational ones [21, 23].

Significant research challenges remain in relation to lessening the distance between human intelligence and AI. These challenges include: (a) Addressing limitations such as the reliance on humanlabelled data, reliance on reward and brittleness in the face of changing context; and (b) extending current abilities in relation to symbol manipulation, reasoning, causal inference, and what we colloquially refer to as common sense [5]. Work is also required that is more specifically geared to bodily animation and activity recognition in virtual environments [9].

3.3 Expansive environment(s)

A term increasingly used as shorthand for complex virtual environments is the metaverse, sometimes described as a three- dimensional virtual social space [25]. Embraced by some of the world's largest platform providers [26], this vision proposes the distinction between digital and physical domains all but disappear. Consequently, expansive environments that consist of photo-realistic and interactive entities (objects and other virtual humans, etc.) is the third element of our framework.

The environment in which interactions with virtual humans take place is determined by their context of use. To-date, most interactions take place in empty spaces (often a white background), which though appropriate in some scenarios, are often used to reduce complexity. This is likely to change in future experiences of increasingly detailed and complex environments. We see two key areas where the state-of-the art is lacking here. First, that environments remain somewhat 'dumb'—where assets generally lack the information (intelligence) to support the behavior (interaction) required for BVHs. Second, networked intelligence may well be required for environments to be generated dynamically (as opposed to pre-programmed) and respond to the emotional and psychological state of users. Third, for effective bodily interaction, haptic technology needs to move beyond the bulky hardware associated with current Virtual Reality (e.g., headset, gloves, etc.). Further work required in this regard: (a) Traditional actuators (e.g., eccentric rotating mass and linear resonant) are evolving to include contactless approaches (e.g., air jet and ultrasonic radiation); and (b), haptic sensors are starting to make their way into everyday clothing. Fourth, we need to better understand the user experience in all of this [13].

4 PROPOSITIONS

The outline framework shown in Figure 1 summarizes the three elements we have presented so far: Appearance, behavior, and environment. To stimulate further debate, we now discuss a set of propositions for each core component of that framework. We posit that, once all elements in the framework are integrated, virtual humans' interactions will become "to most intents and purposes indistinguishable from physical humans" [7, p. 247]. In other words, the vision of a multi-agent system sharing a common environment will become a reality for BVH experiences (cf. [17]).

We propose that:

- The appearance of BVHs must be distinct rather than generic, that is it must require a unique photo-realistic human-like appearance that will change over time in the same way our own appearance is distinct and changes over time.
- The behavior of BVHs must be autonomous rather than scripted, which requires a behavioral logic to support changing appearances and interactions with the environment, that is with objects and other agents.

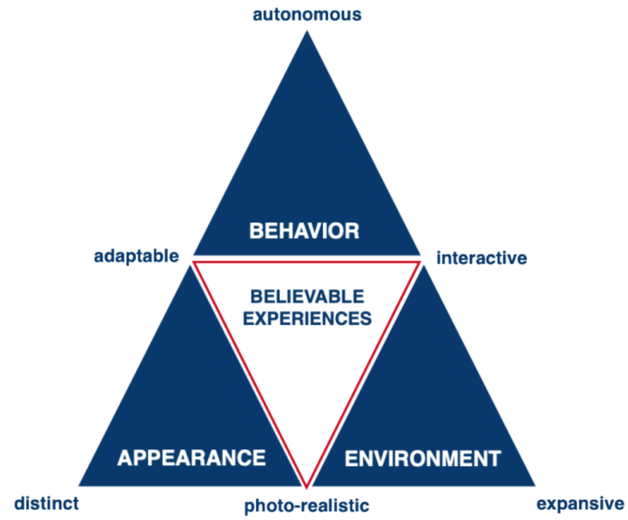

Figure 1: Proposed framework to analyze believable virtual human experiences

- BVHs are embedded in an expansive and dynamic environment that is populated by photo- realistic objects with which virtual humans interact (which is likely to be a necessary pre-condition for the realization of the metaverse).

5 DISCUSSION ON CHALLENGES ARISING

Several key challenges must be addressed to realize this vision. Besides specific technical challenges associated with human-like appearances, autonomous behavior, and complex environments, we believe that the most pressing challenge is the integration of these increasingly sophisticated technologies (highlighted as a red triangle in Figure 1). In the following, we briefly discuss each challenge.

5.1 Human-like appearance

Accepting that it is always possible for creators to 'simplify their assignment,'—e.g., by using techniques such as photo-faces, where a two-dimensional photo of a face is placed onto a generic head model [7]—we remain optimistic that specific technical challenges will be solved. Key, however, is that the appearance of BVHs must be adaptable, that is to correspond with their behavior, and show the same level of photo-realism as their environment (non-trivial challenge).

Moreover, it is not yet clear what level of (technical) sophistication audiences demand and/or respond to. Here, we see key research challenges as including: (a) Providing effective and efficient ways for users to create and manage virtual appearances (including secure ownership for managing and exchanging assets, incl. brand management etc.); (b) understanding and managing the degree to which human distinctiveness and identity is modelled and reflected real-time during interactions and in different contexts; and (c) understanding and managing how BVHs visually adapt to and age within their environment.

5.2 Autonomous behavior

Navigating through and interacting within environments remains a key challenge. Even though conversational AI may not be the most efficient way of communicating information (e.g., adding short pauses when none are required), it shows great potential to ease the interaction with complex technologies; even though performance of dialogue systems varies widely with the same underlying model to process language [11].

Here, we see the key research challenges as including: (a) Enabling expressive ability in real-time and at scale; and to understand the degree to which fine facial motor behavior must match bodily behavior; (b) improving and extending the data that is used to manage responses; (c) developing haptic approaches that tie immersive experiences to less cumbersome devices than at present; (d) developing more human-like cognitive architectures that match the expressive abilities expected from BVHs; (e) improving the conversational effectiveness of BVHs that fulfil specific organizational or social/entertainment roles (e.g., customer service); (f) constraining behavior to that which is appropriate for specific virtual environments (incl. to limit the effect of bias inherent in training data); (g) experimenting with ways in which virtual humans engage users and the roles they play; and (h) evaluating factors to enhance and influence audience engagement [7].

5.3 Expansive environment

A further key challenge of BVH experiences will be to manage interactions within and across environments. As in the physical world, virtual environments are likely to expand over time in scope and in the number of objects and entities that populate them.

Increasing environmental complexity can be reduced by visually highlighting objects a user can interact with and/or by limiting the environment to a manageable number of items and characters. It is likely that these techniques will also have to be employed with BVHs experiences, at least initially. This points to the growing importance of storytelling and the way convincing (believable) narratives unfold. While addressing the compatibility and expansive nature of digital environments relates predominantly to technology and business models, environmental complexity will likely be managed through storytelling; not only to ease the transition towards increasingly complex environments, but also to achieve believability through immersion. Here, we see the key research challenges as including: (a) Better understanding (and developing appropriate research methods for) how we interact in digital spaces via BVHs—the user experience; (b) improving our understanding of the design of digital 'space' and 'place' from conceptual, information, (work) process, social and entertainment perspectives; (c) exploring the degree to which users want to record/own their role (i.e. history) withing the story, (d) embedding intelligence into environmental assets as a means of informing place and space to guide the behavior and contextual intelligence of BVHs; (e) networking this intelligence such that environments can be built on the fly and respond to the emotional and psychological state and/or social need; and (f) exploring the importance of narratives in creating believable and immersive experiences (despite the knowledge that interaction is with a virtual person and not a real one).

5.4 Integration

The technologies needed to facilitate a distinct appearance, autonomous behavior, and expansive environments are wide ranging, as we have previously noted, including: (a) VFX technologies such as volumetric capture, lidar scanning etc.; (b) advanced haptics; (c) more cognitive-focused AI; (d) content/context awareness; (e) machine vision; (f) real-time game engines; (g) immersive content and intelligent asset management; (h) non-linear storytelling; (i) blockchain (including smart contracts and non-fungible tokens for provenance of assets); and (j) data analytics to manage and personalize information.

A particular challenge will be the integration of these various technologies into a coherent user experience. Though not exhaustive, the overarching research point is that frameworks and standards will likely need to be created to ensure that technologies work together effectively and efficiently and personalized to the context of use.

6 CONCLUSION

While the vision we have outlined at the beginning is still some time off, our brief review has shown that the technologies necessary to create BVH experiences already exist.

Consequently, the creation of believable virtual human experiences encompasses the technical challenge of integrating those technologies alongside a creative and social one–the latter because virtual human experiences must have a (narrative) purpose and engage audiences to immerse themselves willingly in the experience. The (inter)subjective nature of the social challenge mandates that significant research is thus required to:

- Identify and prioritize the most urgent use cases.
- Better understand the needs, wants, and preferences of diverse audiences in their context of use.
- Assess the type(s) of visual interface(s) users respond to.

These are some of the priorities we have set for ourselves to evaluate the vision of believable virtual human experiences. We hope that our contribution stimulates debate on these research priorities at the conference.

REFERENCES

[1] Karim Abdel-Malek *et al.*, 2007. Development of the virtual-human Santos □ùëá□ùëÄ. Digital Human Modeling. Vincent G. Duffy (Ed.). Springer, Berlin, 490-499. ISBN: 9783642028083.
[2] Francisco Azuola *et al.*, 1994. Building anthropometry-based virtual human models. In: Proceedings of the IMAGE VII Conference, Tucson, AZ, 54.
[3] Norman I. Badler, 1997. Real-time virtual humans. In: Proceedings of the Fifth Pacific Conference on Computer Graphics and Applications. IEEE, 4-13.
[4] Norman I. Badler, Michael J. Hollick and John P. Granieri, 1993. Real-time control of a virtual human using minimal sensors. Presence: Teleoperators and Virtual Environments, 2, 1 (Feb. 1993), 82-86. DOI: https://doi.org/10.1162/pres.1993.2.1.82.
[5] Yoshua Bengio, Yann Lecun and Geoffrey Hinton, 2021. Deep learning for AI. Communications of the ACM, 64, 7 (Jun, 2021), 58-65. DOI: 10.1145/3448250.
[6] David Pal Boros and Karoly Hercegfi, 2020. Digital human modelling in research and development-a state of the art comparison of software. Human Systems Engineering and Design II. Tareq Ahram *et al.* (Eds.), 543-548. ISBN: 978-3-030-27928-8.
[7] David Burden and Maggi Savin-Baden, 2019. Virtual humans: Today and tomorrow. London: CRC Press.
[8] Ergun Ekici, 2020. What's the difference between a chatbot and conversational AI? Amelia, an IPsoft Company. URL: https://youtu.be/xHAnLceQnT0 (visited on 20/06/2021).

[9] Fuqiang Gu *et al.*, 2021. A Survey on deep learning for human activity recognition. ACM Computing Surveys, 54, 8 (Oct., 2021). DOI: https://doi.org/10.1145/3472290.

[10] Mubbasir Kapadia *et al.*, 2012. What's next? The new era of autonomous virtual humans. Motion in Games. Marcelo Kallmann and Kostas Bekris (Eds.). Springer, Berlin, 170-181. DOI: https://doi.org/10.1007/978-3-642-34710-8_16.

[11] Seokhwan Kim *et al.*, 2021. Overview of the eighth dialog system technology challenge: DSTC8. IEEE/ACM Transactions on Audio, Speech, and Language Processing (May, 2021), 2529-2540. DOI: https://doi.org/10.1109/ TASLP.2021.3078368.

[12] Maria de Kleijn, Mark Siebert and Sarah Huggett, 2019. Knowledge management with digital humanities / digital scholarship. In: Artificial Intelligence: How knowledge is created, transferred and used (IFLA WLIC). Corfu, Greece. URL: http://library.ifla.org/2814/.

[13] Ahmed K$\sqrt{\partial}$se, Aleksei Tepljakov and Eduard Petlenkov, 2021. Intelligent virtual environments with assessment of user experiences. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, volume 12854, 463-373. DOI: https://doi.org/10.1007/978-3-030-87986-0_41.

[14] John E. Laird, 2008. Extending the Soar cognitive architecture. In: Proceedings of the 2008 Conference on Artificial General Intelligence, 171. IOS Press, 224-235.

[15] Hobson Lane, Cole Howard and Hannes Max Hapke, 2019. Natural language processing in action: Understanding, analyzing, and generating text with Python. Manning, Shelter Island, NY. ISBN: 9781617294631.

[16] Antonio Lieto *et al.*, 2018. The role of cognitive architectures in general artificial intelligence. Cognitive Systems Research, 48 (May, 2018), 1-3. DOI: https://doi.org/10.1016/j.cogsys.2017.08.003.

[17] Nadia Magnenat-Thalmann and Daniel Thalmann., 2004. An overview of virtual humans. Handbook of Virtual Humans. N. Magnenat-Thalmann and D. Thalmann (Eds.). John Wiley & Sons, Chichester, 1-25. ISBN: 0-470-02316-3.

[18] John Markoff and Tom Gruber, 2019. A conversation about conversational AI. TEDxBeaconStreet (Jan., 2019). URL: https://youtu.be/yoCwsvIyp9Y (visited on 20/06/2021).

[19] Shervin Minaee *et al.*, 2021. Deep Learning-based text classification: A comprehensive review. ACM Computing Surveys, 54, 3 (Apr., 2021). DOI: https://doi.org/10.1145/3439726.

[20] Jinjie Ni *et al.*, 2021. Recent advances in deep learning-based dialogue systems: A systematic survey. arXiv preprint (2021). URL: https://arxiv-download.xixiaoyao.cn/pdf/2105.04387.pdf (visited on 20/06/2021).

[21] Isabella Panella, Luca Zanotti Fragonara and Antonios Tsourdos, 2021. A deep learning cognitive architecture: Towards a unified theory of cognition. Advances in Intelligent Systems and Computing (AISC), volume 1250, 566-582. DOI: https://doi.org/10.1007/978-3-030-55180-3_42.

[22] Gopika Rajendran and Ojus Thomas Lee, 2020. Virtual character animation based on data-driven motion capture using deep learning technique. In: Proceedings of the 15th International Conference on Industrial and Information Systems (ICIIS 2020). 333-338. DOI: https://doi.org/10.1109/ICIIS51140.2020. 9342693.

[23] Mark Sagar, Mike Seymour and Annette Henderson, 2016. Creating connection with autonomous facial animation. Communications of the ACM, 59, 12 (Dec., 2016), 82-91. DOI: https://doi.org/10.1145/2950041.

[24] Mike Seymour, 2018. What happens when technology has a human face? TEDxSydnesSalon (Dec. 2018). URL: https://youtu.be/FPm3ZDKnS6A (visited on 03/06/2021).

[25] Chris Stokel-Walker, 2022. Welcome to the metaverse. New Scientist, 253, 3368 (Jan., 2022), 39-43. DOI: https://doi.org/10.1016/S0262-4079(22)00018-5.

[26] Ben Thompson, 2021. Metaverses. Stratechery (Aug., 2021). URL: https://stratechery.com/2021/metaverses/ (visited on 04/08/2021).

[27] Charlie C. L. Wang *et al.*, 2003. Virtual human modeling from photographs for garment industry. Computer-Aided Design, 35, 6 (May, 2003), 577-589. DOI: https://doi.org/10.1016/S0010-4485(02)00080-5.

[28] Isaac Wang and Jaime Ruiz, 2021. Examining the use of nonverbal communication in virtual agents. International Journal of Human-Computer Interaction, 37, 17 (Mar 2021), 1648-1673, DOI: 10.1080/10447318.2021.1898851.

[29] YouTube, 2019. How far is too far? | The age of A.I. YouTube Originals (Dec., 2019). URL: https://youtu. be/UwsrzCVZAb8 (visited on 21/06/2021).