To appear in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops.

Virtual Human Coherence and Plausibility – Towards a Validated Scale

David Mal[1,2,*] Erik Wolf[1,*] Nina Döllinger[2,*]
Mario Botsch[3,†] Carolin Wienrich[2,*] Marc Erich Latoschik[1,*]

[1] HCI Group, University of Würzburg, [2] HTS Group, University of Würzburg, [3] Computer Graphics Group, TU Dortmund University

ABSTRACT

Virtual humans contribute to users' state of plausibility in various XR applications. We present the development and preliminary evaluation of a self-assessment questionnaire to quantify virtual human's plausibility in virtual environments based on eleven concise items. A principal component analysis of 650 appraisals collected in an online survey revealed two highly reliable components within the items. We interpret the components as possible factors, i.e., *appearance and behavior plausibility* and *match to the virtual environment*, and propose future work aiming towards a standardized *virtual human plausibility scale* by validating the structure and sensitivity of both sub-components in XR environments.

Index Terms: Human-centered computing—HCI theory, concepts and models; HCI design and evaluation methods;

1 INTRODUCTION

Virtual representations of human beings are the rationale of various relevant Virtual, Augmented, and Mixed Reality (in short, XR) applications that support, for example, blended learning, psychological treatment, or motor rehabilitation. Particularly for the development of virtual environments (VE) leveraging those virtual humans, it is important to consider the perceptive and cognitive processes underlying the individual's XR experiences.

An essential concept in the XR context is the plausibility illusion (Psi). It refers to the illusion of "what is apparently happening is really happening." [5, p. 3553] and is contrasted by the objective concept of coherence, describing a virtual scenario's objective characteristics that make it reasonable or, in certain contexts, also predictable [4]. Latoschik and Wienrich [2] proposed an alternative theoretical model describing XR experiences and effects. The authors adopt the idea of Psi and coherence but argue that "there is no plausibility illusion but merely plausibility", with plausibility being defined "as a state or condition during an XR experience that subjectively results from the evaluation of any information processed by [...] sensory, perceptual, and cognitive layers" [2, pp. 1, 5]. As virtual humans are an essential part of various VEs, we frame their plausibility with respect to the previously introduced model of Latoschik and Wienrich [2] and state that *virtual human plausibility (VHP)* is the subjective feeling of how reasonable and believable a virtual human appears to a user on a sensational, perceptive, and cognitive level. Following this definition, VHP is shaped by users' "genetics or life-long habitude perceptions" [2, p. 6], their experiences with virtual humans, and their expectations regarding virtual humans' appearance and behavior in a certain VE. VHP is the plausibility that arises from the coherence of all (implicit) sensory impressions of the virtual human. Accordingly, *virtual human coherence* is composed of the internal coherence of a virtual human's appearance and behavior, its coherence with the VE, and the coherences between all additional sensory impressions of the virtual human.

*e-mail: firstname.lastname@uni-wuerzburg.de
†e-mail: firstname.lastname@tu-dortmund.de

A standardized tool to measure the impact of virtual human coherence on VHP can be an essential resource for researchers developing relevant VEs with virtual humans. Therefore, we propose a first measure to quantify VHP with a concise self-assessment questionnaire. As a first step in evaluating the questionnaire, we analyze data from an online survey conducted to obtain initial assessments, discuss the results of a principal component analysis and suggest future work to validate the questionnaire in XR environments.

2 THE VIRTUAL HUMAN PLAUSIBILITY QUESTIONNAIRE

We developed eleven concise questions to quantify VHP. The items are based on the perceived plausibility of the virtual humans' appearance and behavior coherence (items 1 to 6) and the virtual humans' coherence with the VE (items 7 to 11). For a standardized use of the VHP items, we used the term *virtual character* instead of *virtual human*. This prevents implicit bias as a reference to *human* would likely impose additional expectations for realism and resemblance to human characteristics that do not necessarily reflect the virtual human's plausibility in the specific VE. Users are instructed to read all statements carefully and tick to what extent the statements apply to the virtual character shown. Each rating is assessed on a Likert scale from 1 (*does not apply at all*) to 7 (*completely applies*).

3 METHOD

In an online study with a repeated measures design, participants got consecutively presented ten different video stimuli in randomized order. As the independent variable, each video showed one out of ten virtual humans animated by recorded short motions in a neutral VE (Fig. 1). While the VE remained unchanged, the virtual humans had either a *rather realistic-looking* (6 stimuli) or an *abstract* (4 stimuli) appearance. As the dependent variable, participants rated each virtual human presented using the VHP questionnaire introduced.

Figure 1: Four examples of virtual humans in the virtual scene.

3.1 Participants

76 students of the University of Würzburg participated and received credit points for attending the experiment. Post-survey exclusion criteria ruled out the data of 11 subjects. In the resulting 65 valid data sets ages ranged from 18 to 53 years ($M = 21.52$, $SD = 4.29$) with 33 female, 31 male, and one nonbinary participant.

3.2 Video Stimuli

We created six rather realistic-looking virtual humans by scanning three female and three male volunteers using a state-of-the-art 3D-reconstruction photogrammetry process [1], striving for a life-like

Table 1: Eleven virtual human plausibility questions and the summary of a multilevel principal component analysis with oblique rotated factor loadings for each item of two transformed principal components (TC). Factor loadings over .5 appear in bold.

Item	TC 1	TC 2
(1) The behavior of the virtual character seemed to be plausible to me.	**-0.76**	-0.06
(2) The appearance of the virtual character seemed to be plausible to me.	**-0.73**	-0.02
(3) The virtual character's behavior matched its appearance.	**-0.77**	-0.03
(4) The behavior and appearance of the virtual character were coherent.	**-0.81**	-0.03
(5) The virtual character behaved as I would expect it to behave.	**-0.86**	0.06
(6) I could predict by the virtual character's appearance how it would behave.	**-0.81**	0.10
(7) The virtual character fit into the virtual environment.	0.01	**-0.89**
(8) The virtual character was a plausible part of the virtual environment.	-0.06	**-0.84**
(9) The appearance of the virtual character and the virtual environment matched.	0.03	**-0.91**
(10) The behavior of the virtual character matched with the virtual environment.	-0.03	**-0.80**
(11) The virtual character behaved in the virtual environment as I would expect it to.	**-0.58**	-0.25

appearance by resembling human features. Four abstract virtual humans based on a wooden mannequin were used to show only very generic anthropomorphic features. Unique, distinguishable variants of the same abstract virtual human were created by coloring the mannequins' upper bodies in green, yellow, blue, or wooden color. To create video stimuli (18 seconds, 30 fps, 1280×720 px) we animated the virtual humans with an 11 seconds idle motion, followed by a short right handed wave motion. Moreover, we added unobtrusive pre-modeled facial expressions, including randomized eye blinks. The technical system was inspired by Wolf et al. [6].

4 Multilevel Principal Component Analysis

A multilevel PCA variation was performed using an R package provided by Rohart et al. [3] to identify the underlying structure or domains of the newly created VHP items. To account for possible dependencies between repeated measures, all ratings of the 65 participants were combined to a resulting data matrix of 650 appraisals, and the PCA was performed on the within-subject deviation matrix instead of the covariance matrix.

4.1 Results

The sampling adequacy for the analysis was verified (KMO = .94) and correlations between items were sufficiently large, $\chi^2(55) = 5269.81$, $p < .001$. By analyzing the scree plot and applying Kaiser's criterion to an initial analysis, we found two suitable principal components (PC) with an eigenvalue above 1.0. PCs were oblique rotated (oblimin), resulting in the first transformed component (TC) explaining 37.18% of the variance and the second TC explaining 27.72% of the variance, with a total 64.89% of the variance being explained by the two TCs. High internal reliability was found for both components with Cronbach's $\alpha > 0.91$. Table 1 shows the loadings for each of the transformed components.

4.2 Interpreting the Components

When interpreting the components as possible sub-factors of VHP, it becomes apparent that each of them is bound to its respective frame of reference. The first component (TC1) consists of items 1 to 6 and 11. The frame of reference for these items seems to be the virtual human itself. Its behavior is judged by its appearance, and its appearance is judged in the context of its behavior. We interpret this component as a sub-factor of VHP, describing the match between a user's individual experiences and expectations with the coherence of a virtual human's appearance and behavior. Thus, we name it virtual human *Appearance and Behavior Plausibility (ABP)*. The second component (TC2) consists of items 7 to 10. Here, the VE appears to serve as a reference frame for VHP. The items reflect a user's subjective perception of a virtual human's coherence with the VE. A resulting factor might contribute to the perceived understanding of a virtual human's *Match to the Virtual Environment (MVE)*. The items' loadings indicate a clear separation between ABP and the VE-related items of the MVE component. A special case comes with item 11. It covers both the match of the behavior and the environment and the expectations on the virtual human's behavior itself. The loadings reflect the item's focus on ABP over the environmental match (MVE). Following our interpretation, the identified components fit our predefined reference frames for virtual human coherence, namely *virtual human appearance and behavior coherence* for ABP and *virtual human environment coherence* for MVE. For both factors, we expect that a higher score would generally lead to a higher VHP and consequently to a higher overall plausibility.

5 Conclusion

In this work, we present the first version of a VHP questionnaire with two sub-scales, concerning (1) the virtual human's *Appearance and Behavior Plausibility* and (2) the virtual human's *Match to the Virtual Environment*. The eleven concise self-report questions (Table 1) were developed based on introduced theory and evaluated in an online study in which participants rated two types of virtual humans (rather realistic-looking/abstract). Principal component analysis indicated two sub-scales with high reliability, supporting the underlying theoretical structure of *virtual human coherence*. Future work is suggested to aim towards a standardized and robust *virtual human plausibility scale* by further validating the structure and sensitivity of both found sub-components in XR environments.

Acknowledgments

This research has been funded by the German Federal Ministry of Education and Research in the project ViTraS (16SV8219).

References

[1] J. Achenbach, T. Waltemate, M. E. Latoschik, and M. Botsch. Fast generation of realistic virtual humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, New York, NY, USA, 2017. Association for Computing Machinery.

[2] M. E. Latoschik and C. Wienrich. Coherence and plausibility, not presence?! Pivotal conditions for XR experiences and effects, a novel model. *arXiv*, (2104.04846), 2021.

[3] F. Rohart, B. Gautier, A. Singh, and L. C. Kim-Anh. mixomics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology*, 13(11):e1005752, 2017.

[4] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton. A survey of presence and related concepts. *ACM Computing Surveys*, 50(6):96:1–96:39, 2017.

[5] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.

[6] E. Wolf, N. Döllinger, D. Mal, C. Wienrich, M. Botsch, and M. E. Latoschik. Body Weight Perception of Females using Photorealistic Avatars in Virtual and Augmented Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2020.