Evaluation of Omnipresent Virtual Agents Embedded as Temporarily Required Assistants in Immersive Environments

Andrea Bönsch[1,2,*] Jan Hoffmann[1] Jonathan Wendt[1,2,*] Torsten W. Kuhlen[1,2,*]

[1]Visual Computing Institute, RWTH Aachen University, Germany [2]JARA-HPC, Aachen, Germany

Abstract

When designing the behavior of embodied, computer-controlled, human-like virtual agents (VA) serving as temporarily required assistants in virtual reality applications, two linked factors have to be considered: the time the VA is visible in the scene, defined as presence time (PT), and the time till the VA is actually available for support on a user's calling, defined as approaching time (AT).

Complementing a previous research on behaviors with a low VA's PT, we present the results of a controlled within-subjects study investigating behaviors by which the VA is always visible, i.e., behaviors with a high PT. The two behaviors affecting the AT tested are: *following*, a design in which the VA is omnipresent and constantly follows the users, and *busy*, a design in which the VA is self-reliantly spending time nearby the users and approaches them only if explicitly asked for. The results indicate that subjects prefer the *following* VA, a behavior which also leads to slightly lower execution times compared to *busy*.

Index Terms: H.5.2 [User Interfaces]: Evaluation/Methodology—User-Centered Design I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality;

1 Introduction

Direct social interactions with computer-controlled, intelligent and conversational virtual agents (VAs) are an integral part of various virtual reality (VR) applications. The majority of use cases are thereby training applications. Here, VAs either represent instructors (e.g., [4]), peers (e.g., [11]), or training partners like negotiation partners (e.g., [5]), staff at a shared workplace (e.g., [9]), or pupils (e.g., [7]) enabling users to learn and improve special skills, e.g., in the domain of social or motor skills. In addition, VAs may also function as scene guides (e.g., [10]) or as interlocutors (e.g., [8]) imparting knowledge about certain scene locations or answering topic-specific questions.

In such agent-based systems, the VA is typically designed as an omnipresent companion, who is in an almost permanent social interaction with the user. However, there are also use cases in which a VA is only required temporarily [2]. Examples are virtual assistants who only support if there is a need for assistance, e.g., when users have a specific question or when they make a mistake in a task execution. This setting raises the need for a suitable VA's behavior in the spare time, defined as the time in which no social interaction takes place and users work on their own, as well as in the moment in which the users try to establish contact with the VA.

As discussed by Bönsch et al. [2], two linked factors should be considered when designing such temporarily required virtual assistants, illustrated in Figure 1: the *presence time (PT)*, defining how long a VA is visible, and the fallback time, renamed to *approaching time (AT)* in [3], defined as the time span between the user's call for support and the VA's actual availability.

In a first within-subjects study, three behaviors affecting the AT, while keeping a low PT (cp. Fig. 1, column 1), were evaluated,

*e-mail: {boensch|wendt|kuhlen}@vr.rwth-aachen.de

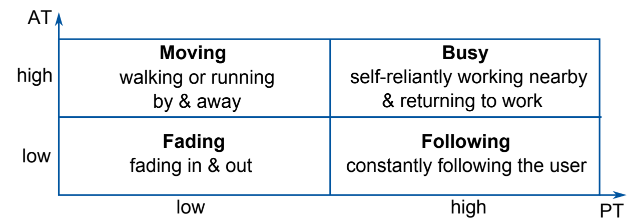


Figure 1: Behavior options of temporarily required VAs distinguished by the two linked factors presence time (PT) and approaching time (AT).

namely fading as an instant appearing and disappearing of the VA, as well as walking and running as two more realistic and human-like behaviors [3]. The insights gained did not reveal a clear preference towards either a realistic or a non-realistic approaching-behavior. Instead the study results indicated that more research is required on a suitable trade-off between PT, AT and a realistic VA's behavior.

To complement the first study, the work presented here focuses on column 2 of Figure 1, i.e., behaviors with a high PT. By conducting a second within-subjects study, we investigate whether users prefer an omnipresent VA, who follows them wherever they go, or a VA who spends time self-reliantly in the user's vicinity and approaches the user only if explicitly asked for.

The remainder of this paper is structured as follows: We present the study design and hypotheses in Section 2, summarize the main results in Section 3, discuss them in Section 4 and give a short summary and outlook in Section 5.

2 User Study

In this work, we investigate the users' preference on two different ATs for temporarily required virtual assistants, while keeping the PT constant by presenting the VA permanently within the scene. To gain the relevant insights, we conducted a within-subjects user study within a CAVE. We designed a basic search task, in which subjects had to find an item in a scene, pick it up, and hand it over to the VA. This task was repeated several times per scene. One scene was a small-scale environment fitted to the CAVE's footprint, requiring only natural walking of the subjects to cross the scene. In contrast, the second scene was a large-scale environment, enriched by moving into a pointing direction as artificial navigation technique. Two behaviors for the VA are tested: a constant following behavior (a_{follow}) and a behavior in which the VA self-reliantly spends time in the user's vicinity while approaching the subject only on request and returning to the previous activity after receiving the item (a_{busy}).

We expected the following two hypotheses to be confirmed:

H1 *a_{follow} results in a faster task execution.*
As a_{follow} ensures an instant and unhindered support, we expect lower task execution times compared to a_{busy}.

H2 *a_{busy} is preferred over a_{follow}.*
Although subjects need to tolerate a higher AT during a_{busy}, we expect them to favor the self-reliantly occupied spare times of the VA over its permanent presence in close proximity. Being watched the entire time by the VA in a_{follow} may annoy and distract the user, leading to an too unsettling experience.

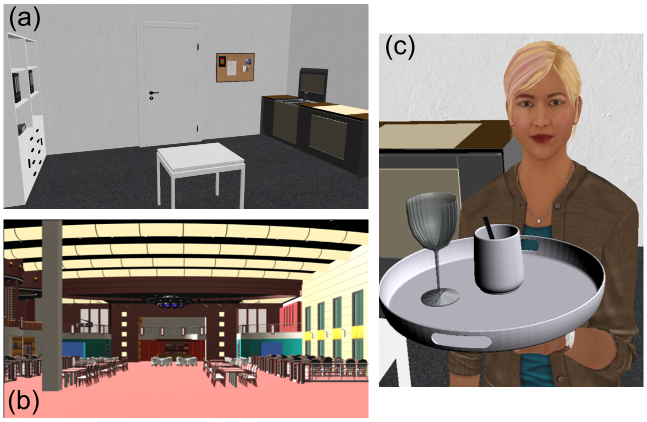


Figure 2: As study scenes a small-scale kitchen (a) and a large-scale restaurant (b) were modeled. A female VA was embedded as assistant collecting the cups and glasses found by the subjects on a plate (c).

2.1 Apparatus

The study was conducted in a five-sided CAVE with the dimensions $5.25m \times 5.25m \times 3.30m$ $(w \times d \times h)$, while the four walls and the floor provide a $360°$ horizontal field of regard. The subjects were equipped with stereo glasses and an ART Flystick 2 for navigation and scene interaction, both tracked at 60 Hz. A microphone array and two security cameras mounted in the CAVE's ceiling furthermore enabled the experimenter to observe the fully immersed subject without being noticed.

2.2 Virtual Environments and Task

Two scenarios have been used in the study: a small-scale, sparsely furnished kitchen ($s_{kitchen}$) fitted to the CAVE's footprint, shown in Figure 2(a) and a large-scale restaurant ($s_{restaurant}$), shown in Figure 2(b).

A female agent (see Fig. 2(c)) is introduced as virtual assistant Rachel whose task is the subject's support. The character model is taken from and animated by means of SmartBody [12].

Subjects had to fulfill the same task with different numbers of iterations depending on the actual virtual environment: tidying up the scene. Therefore, they had to find an item, either a cup or a glass (see Fig. 2(c)), by means of navigating through the scene, approach it closely to pick it up by a ray-casting-based point-and-click metaphor, and hand it over to Rachel, who collects all found items, by holding the item over her plate. For all interactions the Flystick was used. In $s_{kitchen}$ five items had to be found, in $s_{restaurant}$ the item number was set to twenty. In case Rachel was showing the behavior a_{busy}, the subjects had to call her by means of a designated button on the Flystick.

2.3 Virtual Assistant's Behaviors

We designed two behaviors for the virtual assistant: a_{follow} and a_{busy}.

During a_{follow}, the VA followed the subjects constantly while staying in their backs and thus out-of-sight to avoid occluding the scene. During the subjects' own locomotion, the VA followed them while respecting the subject's personal space by maintaining a distance of approximately $100cm$ (45 to $120cm$ are considered as personal space based on [6]). In case subjects collided with scene geometry, the VA chose a different path avoiding the collision. After subjects stopped, the VA approached them till she is positioned approximately at arm's length away facilitating the expected handing over of an item by the subject. After receiving the item, the VA returned to her following behavior when the user exceeded a distance threshold of $100cm$.

a_{busy} is an extension of a_{follow} where the VA autonomously moved around in the scene unless being called for. When subjects called for the VA, a simple wayfinding algorithm was used to navigate the VA to the subjects while avoiding collisions with the scene geometry. Comparable to a_{follow}, the VA stopped at arm's length away and returned to her idle behavior once the user exceeds the distance threshold of $100cm$. While idling, the VA self-reliantly occupied the spare time in neighboring scene areas of the subjects' current location. For $s_{restaurant}$ we limited the neighboring areas to the same restaurant's quarter in which the subjects are currently searching for items. This design was supposed to result in comparable ATs in $s_{restaurant}$ independent of the subjects' positions in the scene, avoiding extreme ATs in case the VA had to cross the complete scene. The VA spent the spare time by looking at certain, pre-defined objects and locations of interest in the respective scene, e.g., the board in the kitchen scene (see Fig. 2(a)). Per location, the VA spent a few seconds watching, before a new location out of a manually pre-defined location set was assigned and the VA approaches the new location. The idling was immediately terminated on a subject's call for support.

Table 1: Items of the SPS by Bailenson et al. [1].

No.	Item
1.	I perceive that I am in the presence of another person in the room with me.
2.	I feel that the person is watching me and is aware of my presence.
3.	The thought that the person is not a real person crosses my mind often.
4.	The person appears to be sentient (conscious and alive) to me.
5.	I perceive the person as being only a computerized image, not as a real person.

2.4 Experimental Design and Data Collection

We designed a within-subjects study with two independent variables: (a) the scene ($s_{kitchen}$, $s_{restaurant}$) and (b) the VA's behavior (a_{follow}, a_{busy}), resulting in four treatments (2×2). In order to avoid biases, we randomized the treatment order. Thereby, subjects first experienced one of the two VA's behaviors in both scenes, before moving on to the next behavior. The scene order per behavior was also randomized.

We gathered the following data to evaluate our hypotheses: Per treatment, subjects were asked to rate their perceived social presence of the VA by means of the Social Presence Survey (SPS, see Tab. 1) [1]. After experiencing all four treatments, subjects were asked to rate their perceived level of presence by means of the SUS presence questionnaire [13]. Furthermore, we added some study- and task-specific questions, e.g., the subjects were asked to state which of the two behaviors they preferred. In addition, a semi-structured interview in the end revealed more insight. Besides this individual feedback, we measured the execution times per treatment.

2.5 Procedure

On entering our lab, subjects were informed about the general study procedure by means of written instructions. After giving their informed consent, they filled out a demographic questionnaire and entered the CAVE to familiarize themselves with both scenes and the virtual assistant introduced as virtual agent named Rachel. Then, the study began. The execution was divided into two blocks, one per behavior. In each block the subject had to gather all items in both scenes, resulting in two runs. The behavior blocks as well as the order of the scenes were randomized to avoid biases. Per run, the items to be collected were randomly distributed over the scene, to ensure subjects had to search them. Each run was followed by a set of questions, which had to be answered inside the CAVE by using the Flystick. After leaving the CAVE, subjects were asked to fill out a final questionnaire, and a short semi-structured interview completed the study.

2.6 Subjects

21 volunteers from the computer-science department participated in our study (3 ♀, 19 ♂, ages M=29.09, SD=7.24). All had normal or corrected-to-normal vision, normal motor skills, and were naïve to the purpose of the study. 7 subjects never experienced a VR setting before and only 9 subjects interacted with a VA in a VR setting before.

Table 2: Subjects' SPS score per treatment; for SPS items see [1].

		M	SD	sig.	p
$s_{kitchen}$	a_{follow}	17.57	3.34	}+	.06
	a_{busy}	15.33	5.20		
$s_{restaurant}$	a_{follow}	17.52	4.19	}$*$	.02
	a_{busy}	15.29	5.73		

$*$ significant at .05 level, $+$ non-significant trend at .1 level

3 Results and Analysis

For our own questions, we used either a 7-point scale (1 to 7) or concrete response options (the two behaviors), from which exactly one had to be chosen. For the complementing standardized questionnaires, we used the proposed 7-point Likert scale for SUS (1 to 7) and an adapted 6-point Likert scale for SPS (1=strongly disagree to 6=strongly agree). Although the proposed scale for SPS is a 7-point Likert scale (-3 to 3) [1], we decided to use an even-point scale instead to force subjects to either choose a positively or negatively keyed item in order to get a stronger SPS score.

For evaluation, a significance level of 0.05 was used for all tests.

Besides adapting the SPS scale as described before, we also modified the computation of the score. While Bailenson et al. sum up the five individual SPS ratings [1], we decided to invert the ratings for item 3 and 5 during the summation (see Tab. 1 for item clarification). This is due to the fact that a high value, so an agreement, for items 1, 2 and 4 indicates that the VA is perceived as a person, while a high value for items 3 and 5 indicates that the VA is perceived as computer interface. Based on our two adaptions, the SPS can take values from 5 to 30. The SPS scores are shown in Table 2 with M denoting the mean and SD denoting the standard deviation. By paired-sampled t-test, no significant difference between the four treatments could be revealed with respect to the behavior. However, for $s_{restaurant}$ a significant difference between both behaviors was revealed, and a non-significant trend for $s_{kitchen}$.

To get a further indication whether the VA was perceived as human companion, we asked the subjects whether they prefer to collide with the VA or with scene geometry in case of an indispensable collision (see row 1 in Tab. 3). A 7-point Likert scale from 1=strongly disagree to 7=strongly agree was used. A one-sample t-test revealed that the mean is significantly less than 4 at .05 level.

We also asked our subjects which behavior of the VA they preferred. While only four subjects preferred a_{busy}, the remaining seventeen subjects voted for a_{follow}. A one-sample binomial test revealed a significant difference between both ratings (p=.004).

With respect to a_{busy}, we asked our participants to rate the perceived integration of the VA into the virtual environment (VE). The details are given in row 2 of Table 3. A one-sample t-test revealed no significance that the actual mean is greater than 4.

The subjects rated the feeling of being present in the VE as reasonably high. An average SUS score of M=4.12 (SD=0.43) was reported.

Finally, we evaluated the execution times per treatment and compared it with respect to the scenes. The numbers are listed in Table 4. Although the average times show larger execution times for a_{busy}, two-sided, paired-sampled t-tests per scene did not reveal any significances. We further explored the large deviations in the task execution. While we found no explanation for $s_{restaurant}$, single outliers caused the large deviations in $s_{kitchen}$. However, no link between the outliers and the respective subjects could be revealed. Correcting for the outliers, the following means and standard deviations are found: a_{follow} with M=50.98, SD=13.15 for 20 samples and a_{busy} with M=54.92, SD=13.02 for 18 samples.

4 Discussion

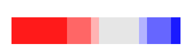
Throughout the study, the interaction between the user and the VA stayed the same. To this end, preferences of the subject can be traced back to the VA's behavior.

The visual appearance of the embedded VA is human-like. However, realizing adequate human motions is still challenging and results frequently in unrealistic or robot-like movements. This mismatch influences the perceived social presence and thus, whether the VA is accepted and perceived as a human companion. Our study results indicate, that subjects were aware of the VA's non-human nature, while still dealing with the VA as a human. They, e.g., preferred colliding with the scenery over a collision with the VA and even proposed small extensions for the overall VA's behavior and realization: One third of the subjects asked for short utterances on the VA's approaching and departure as well as for footstep sounds as positional VA feedback within the VE. Due to the lack of the sound feedback, four subjects even stated that they got frightened when they suddenly realized the VA was standing close by. Besides the mentioned footstep sounds, enriching the scene by more sound cues, e.g., breathing sounds or rustling of the VA's clothes, might further improve the setting. Regarding the interpersonal distance, sticking to the personal space as described in Section 2.4 seemed to be a good choice: While five subjects stated, that they perceived the interpersonal distance as too small, five other subjects even wished that the VA would have approached them a bit more.

With respect to the VA's locomotion twelve subjects would have preferred faster movements to reduce the AT, especially in $s_{restaurant}$. Due to a miss-calibration of the artificial navigation technique used by the subjects, they were able to change their position within the VE faster compared to the VA. Thus, the AT times were higher than originally intended, as the VA had to overcome larger distances. However, this shortcoming impacted both tested behaviors equally, thus the miss-calibration did not turn out to be a pitfall for the study. Instead, all subjects adjusted to the speed shortcoming quickly and approached the VA by themselves to reduce the AT. Comparing the task execution times per scene for both behaviors, the results reveal slightly higher times for a_{busy} compared to a_{follow}. However, as the difference is not significant, we cannot confirm **H1**.

When asked which behavior the subjects preferred, they significantly favored a_{follow}. Thus, **H2** cannot be confirmed. We explain this unexpected finding as follows: Based on the SPS scores, a_{follow} was rated more human-like compared to a_{busy} and is thus in favor for a human-like virtual assistant. Especially for $s_{restaurant}$, the difference in perception was significant, while only a non-significant trend could be revealed for $s_{kitchen}$. We assume that the rather short time (cp. treatment execution times, Tab. 4) the subjects spent with the VA in $s_{kitchen}$ was too short for a significant impact of the VA's follow-behavior on the perceived social presence. Perceiving the VA as less human-like in a_{busy}, might be due to the chosen spare time activities of concentrating on certain pre-defined objects and locations of interest in the scene. Our goal was to design a behavior in which the VA should avoid focusing on the user in her spare time, while choosing an activity which can be terminated very suddenly on a user's explicit call for support. How-

Table 3: Results of subjects' preference ratings.

	answer frequencies	M	SD	p	sig.
If there is no way to avoid a collision by means of physical walking, I prefer to collide with Rachel instead of – at least waist-high – scene geometry.		3.14	1.98	.03	$*$
When Rachel focused scene objects like the clock or the board, I felt she was integrated adequately into the VE.		4.33	1.32	.14	

strongly disagree 1 2 3 4 5 6 7 strongly agree

$*$ significant at .05 level

Table 4: Task execution times per treatment in seconds.

		M	SD	p
$s_{kitchen}$	a_{follow}	58.69 s	36.78 s	.57
	a_{busy}	64.70 s	28.15 s	
$s_{restaurant}$	a_{follow}	405.16 s	125.77 s	.33
	a_{busy}	452.06 s	150.58 s	

ever, the subjects' ratings regarding an adequate integration of the VA in her spare time revealed only a neutral overall score, indicating our design as acceptable, however, improvable. Furthermore, three subjects stated, that they hoped for an ongoing support of the VA in the spare time, i.e., the VA was supposed to autonomously search and pick up the items if not required by the subject. Therefore, we recommend improving the VA's spare time activities for ongoing works.

Another interesting insight was gained during the semi-structured interviews. While seven subjects asked explicitly for a more human-like assistant, three subjects characterized the VA as a tool and thus voted for practicability and lower ATs, even to the detriment of realistic human behavior. Thus, no real preference towards either the VA's realism with respect to human-like behavior or to quick and unhindered support was revealed. This supports the statement of Bönsch et al., that more research is required to find a suitable trade-off between PT, AT and the VA's realism [3]. Interestingly, this finding was revealed more clearly in the previous study. We assume this is caused by the different study tasks. While our VA had a reasonable task of collecting the found items, the VA in [3] was purely embedded as interlocutor answering questions. This functionality could have been also realized by simple tooling such as a natural language user interface known, e.g., from Apple's Siri or Amazon's Alexa.

5 Conclusion

In this work, we presented the evaluation of two behaviors (a_{busy} and a_{follow}) for temporarily required assistants. The results indicate that a_{follow} was clearly preferred over a_{busy} and that the task execution times of a_{busy} were only slightly larger due to higher ATs. Further insights additionally supported the finding of a first study on assistants' behaviors [3], that a suitable trade-off between the VA's realism on one side and the PT and AT on the other side need to be found.

For future work, suitable tasks need to be identified in order to carefully evaluate the single parameters influencing the acceptance of a human-like companion embedded as temporarily required assistant.

References

[1] J. N. Bailenson, J. Blascovich, A. Beall, and J. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence*, 10(6):583–598, 2001.

[2] A. Bönsch, T. Vierjahn, and T. W. Kuhlen. Evaluating Presence Strategies of Temporarily Required Virtual Assistants. In *Intelligent Virtual Agents: 12th International Conference, IVA 2016. Proceedings*, pages 387–391. 2016.

[3] A. Bönsch, T. Vierjahn, and T. W. Kuhlen. Evaluation of Approaching-Strategies of Temporarily Required Virtual Assistants in Immersive Environments. In *IEEE Symposium on 3D User Interfaces*, 2017.

[4] I. de Kok, J. Hough, F. Hülsmann, M. Botsch, D. Schlangen, and S. Kopp. A Multimodal System for Real-Time Action Instruction in Motor Skill Learning. In *Proceedings of the 2015 ACM on the International Conference on Multimodal Interaction*, pages 355–362, 2015.

[5] J. Gratch, D. DeVault, and G. Lucas. The Benefits of Virtual Humans for Teaching Negotiation. In *Intelligent Virtual Agents: 12th International Conference, IVA 2016. Proceedings*, pages 283–294. sep 2016.

[6] E. Hall. *The Hidden Dimension: Man's Use of Space in Public and Private.* The Bodley Head Ltd, 1966.

[7] J.-L. Lugrin, M. E. Latoschik, M. Habel, D. Roth, S. Christian, and S. Grafe. Breaking Bad Behaviors: A New Tool for Learning Classroom Management Using Virtual Reality. *Frontiers in ICT*, 3, 2016.

[8] S. McGlashan. Speech Interfaces to Virtual Reality. In *Proceedings of the 2nd International Workshop on Military Applications of Synthetic Environments and VR*, 1995.

[9] A. Robb, A. Cordar, S. Lampotang, C. White, A. Wendling, and B. Lok. Teaming Up with Virtual Humans: How Other People Change Our Perceptions of and Behavior with Virtual Teammates. *IEEE Transactions on Visualization and Computer Graphics*, 21(4):511–519, 2015.

[10] A. Roque, D. Jan, M. Core, and D. Traum. Using Virtual Tour Behavior to Build Dialogue Models for Training Review. In *Intelligent Virtual Agents: 10th International Conference, IVA 2011. Proceedings*, pages 100–105, 2011.

[11] K. Ryokai, C. Vaucelle, and J. Cassell. Virtual Peers as Partners in Storytelling and Literacy Learning. *Journal of Computer Assisted Learning*, 19(2):195–208, 2003.

[12] A. Shapiro. *Motion in Games*, volume 7060, chapter Building a Character Animation System, pages 98–109. 2011.

[13] M. Usoh, E. Catena, S. Arman, and M. Slater. Using Presence Questionnaires in Reality. *Presence*, 9(5):497–503, 2000.