Being Guided or Having Exploratory Freedom: User Preferences of a Virtual Agent's Behavior in a Museum

Andrea Bönsch
Visual Computing Institute,
RWTH Aachen University, Germany
boensch@vr.rwth-aachen.de

David Hashem
Department of Computer Science,
RWTH Aachen University, Germany

Jonathan Ehret
Torsten W. Kuhlen
Visual Computing Institute,
RWTH Aachen University, Germany

ABSTRACT

A virtual guide in an immersive virtual environment allows users a structured experience without missing critical information. However, although being in an interactive medium, the user is only a passive listener, while the embodied conversational agent (ECA) fulfills the active roles of wayfinding and conveying knowledge. Thus, we investigated for the use case of a virtual museum, whether users prefer a virtual guide or a free exploration accompanied by an ECA who imparts the same information compared to the guide. Results of a small within-subjects study with a head-mounted display are given and discussed, resulting in the idea of combining benefits of both conditions for a higher user acceptance. Furthermore, the study indicated the feasibility of the carefully designed scene and ECA's appearance.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; **User studies**.

KEYWORDS

virtual agents, embodied conversational agents, museum, guiding, free exploration, enjoyment, virtual reality

ACM Reference Format:

Andrea Bönsch, David Hashem, Jonathan Ehret, and Torsten W. Kuhlen. 2021. Being Guided or Having Exploratory Freedom: User Preferences of a Virtual Agent's Behavior in a Museum. In *21th ACM International Conference on Intelligent Virtual Agents (IVA '21), September 14–17, 2021, Virtual Event, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3472306.3478339

1 INTRODUCTION

Immersive virtual environments (IVEs) are increasingly often used as learning environments. Examples are training simulations to improve motor (e.g., [16]), social (e.g., [22]), or professional skills (e.g., [37]), but also social or cultural applications. These cover simulations of, e.g., reconstructed places with a high cultural or historical significance (e.g., [4]), cities for VR tourism (e.g., [47]), or digital museums (e.g., [39]), the last being our targeted use case.

Thanks to the large availability of consumer head-mounted displays (HMDs), this immersive content is accessible to many users independently of time and place. However, a narrative layer, providing users with additional information on the exhibits visited, is often missing [26, 27]. To this end, virtual guides (e.g., [13, 30, 44]), presented as anthropomorphic, computer-controlled embodied conversational agents (ECAs), are often embedded as pedagogical agents. Furthermore, these ECAs typically guide the user through the scene, allowing a structured experience [15] while avoiding that the user gets lost in complex scenes or misses critical information [36]. To this end, the user can fully concentrate on gaining knowledge while the ECA takes care of wayfinding, reducing the user's cognitive load.

The question arising, however, is whether being a passive receiver is the best choice for virtual reality (VR) as an interactive medium. Research indicates that active exploration with exploratory freedom increases the sense of personal agency, resulting eventually in higher user investment [11]. One reason might be that users can customize the experience to their individual needs and interest [11].

Our main contribution is an initial exploration of how much independence is preferred by users in a virtual museum. Therefore, we designed two different behaviors for an ECA, representing a member of the museum staff. The behavioral condition C_{guide} simulates a virtual guide, who leads the user on a predefined path through the museum. Condition $C_{companion}$ simulates a virtual companion, who allows the user a free, however, accompanied exploration of the space. Whenever the user shows interest in a special exhibit, the companion provides the same information in the identical way as in C_{guide}. In a pilot study, we investigated which condition is preferred in terms of entertainment, enjoyment, and comfort.

Although our main focus is the ECA's locomotion behavior (e.g., walking speed, proxemics) as a necessity for scene exploration, we also actively considered various design aspects for the ECA as well as the virtual museum and discuss those in the paper. Thus, a second contribution is showing the feasibility of our chosen designs.

Investigating the impact of the ECA's behavioral conditions on the user's spatial and object memory or learning outcome will be done in a later step[1]. We also omitted the simulation of other staff or visitors in this communal space to provide a clean testbed without distractions and external influences for this initial evaluation.

2 RELATED WORK

Using virtual guides in real life to impart knowledge is a common strategy utilizing, e.g., static display-based virtual characters in museums (e.g., [33, 46]) or mobile bots in historic places (e.g., [31]). In addition, using virtual guides or pedagogical agents represented by ECAs in IVEs becomes increasingly common (e.g., [19, 28, 35]). Following the Media Equation Theory [42], these life-like characters are conceived as veritable social actors. Thus, people build relationships with them and react socially [25]. To this end, when designing an

[1] While passive users have a better recognition memory for objects [1], eventually transferable to an increased learning rate, the impact on spatial memory is inconclusive (compare, e.g., [15] and [12])

IVA '21, September 14–17, 2021, Virtual Event, Japan

ECA, a user's expectations of a plausible and realistic human-like attitude and behavior have to be met. For our contemplated mobile ECA various aspects, thus, need to be considered:

In terms of joint locomotion, human-aware navigation is essential, meeting the requirements of comfort, naturalness, and sociability [24, 34]. This also includes respecting the user's personal space (PS) to avoid discomfort (e.g., [2, 8, 23]). During locomotion, footstep sound for the ECA as positional feedback for the users is preferred [7]. Additionally, variations in frequently used trajectories as well as in repeated utterances support perceiving an ECA as more life-like [9].

In terms of conveying information, research showed that natural language has good content delivery [13]. It is recommended to present knowledge in the form of stories [26–28, 35] or anecdotes [32] through an ECA with a personality [3]. There are, e.g., indications that a cheerful guide outperforms a serious one in enjoyment and knowledge transfer [49]. Furthermore, an ECA with a personality becomes more believable, and the interaction feels more natural [3].

In terms of attention guiding, users expect a guide to use gestures and eye contact while standing close to the respective exhibits in real life [32]. Thus, joint attention in IVEs is typically established verbally in combination with gazing and gesturing [19]. In particular, gazing is indispensable for shared objects [29], as it supports all interactants in being mutually aware that they share attention [41]. To support gazing, a suitable formation between the guide and the visitors should be established and maintained [50], allowing the ECA to face the exhibit as well as the listeners. If the attention of a listener wanders, pausing and restarting the current utterance of the ECA is a suitable way to re-attract the user's attention [21]. A user's attention can also wander on the way to the next area of interest, resulting in not following the guide anymore. In those cases, the guide should show situation-aware behavior, i.e., slowing down and waiting for a user who stopped to inspect an object of interest or catching up with a user who departs from the planned tour [40]. Therefore, it is crucial to anticipate the user's intent. While a multitude of behavioral actions can be taken into account [6], gazing and interpersonal distance are essential factors here.

3 USE CASE VIRTUAL MUSEUM

The objective of our work was to investigate whether a classical guided tour (C_{guide}) or a free, however, accompanied exploration ($C_{companion}$) of a virtual museum is preferred. We expected the following hypotheses to be confirmed:

- **H1** The accompanied exploration of a virtual museum is superior to a classical guided tour in terms of enjoyment and comfort in an educational context.
- **H2** The perceived (social) presence will be qualitatively equal in both conditions.

3.1 Equipment & User Navigation

For our study (Sec. 5), we used an HTC Vive, tracked at 90Hz in an area of $3.0m \times 3.0m$ ($w \times d$) using two tripod-mounted SteamVR Base Stations. Subjects brought their own headphones for sound and plugged into the HMD directly. We used one Vive controller for navigation, realized through gaze-based steering. Subjects controlled their speed linearly via the controller's touchpad while asked to stay in place. Informal testing indicated a speed range of 0 to $2.5m/s$ as appropriate.

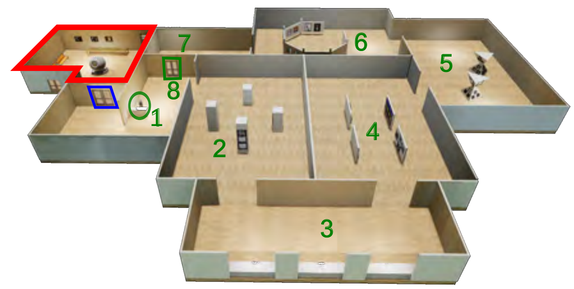

Figure 1: Our museum consists of a training room (red), separated by doors (blue) which open on finishing the training, and the main area. On meeting the ECA (green ellipse, location 1), the exploration starts. After visiting all exhibition rooms (2, 4, 5, 6) and interim hallways (3, 7), the doors at 8 open, triggering the farewell.

3.2 Immersive Virtual Environment

Virtual Museum. Our museum is in a deliberately simple and classical design, depicted in Figure 1. Its exhibition features the history of internet memes focusing on image macros, defined as a picture or artwork with some text superimposed. It is divided into four exhibition rooms (numbers 2, 4, 5, and 6 in Fig. 1), each of which is dedicated to a specific era of internet culture. The visiting order of the rooms is given by the museum's architectural design, going from locations 1 to 8, with 3 and 7 being empty interim hallways. The exploration thus ends in front of the first exhibition room, allowing the user to easily revisit the museum on his or her own after the farewell with the ECA. To avoid touring the museum the other way around in $C_{companion}$, the doors at 8 are closed till all previous locations have been visited. The room outlined in red (see Fig. 1) is purely dedicated to familiarizing the user with the navigation control before the museum's exploration and is thus part of the user study.

The memes[2] are displayed as follows: room 2 presents them on pillars, room 4 on large floating boards, room 5 on hourglass-shaped sculptures, and room 6 comprises a circular arrangement of panels. These varying layouts, first, allow an easy distinction of the individual exhibition rooms [36], while making the layout more interesting. Second, various orders to look at the single memes per room are given, inviting users to freely explore the room by themselves.

General Design Requirements. Although our two behavioral conditions are evaluated for the use case of a virtual museum, they can be applied to various architectural scenarios to be explored by users. Thereby, some general design criteria have to be met: to facilitate scene exploration in open-world scenarios, different design patterns are proposed (e.g., [36]). One required in the context of our work is dividing the IVE into different zones with respective themes [36]. This structures the scene visually as well as from the narrative flow as the ECA introduces each area specifically. Another recommendation is using virtual guides [36] responsible for the global wayfinding, which reflects our C_{guide} condition. To be able to easily configure and alter guided tours through a given environment or specify the companion's behavior at an area of interest, design principles of annotated environments are utilized [17]. These annotations, e.g.,

[2] https://knowyourmeme.com/

encode the positions to be visited and their respective order (used for guiding only) and the utterances and explanations given when reaching the spot. We furthermore assume that different *focal points* on an object of interest are stored, to which the ECA looks when explaining, as well as locations from which the ECA conveys the knowledge. Finally, to pause and restart utterances for attention re-attraction [21], timestamps encoding the start of each word have to be available for the sound samples. When being interrupted, the ECA can then resume the explanation starting with the last word.

3.3 ECA

Subjects meet our ECA Kate at her counter in the entrance hallway (location 1 in Fig. 1). She is depicted as a young woman in smart casual attire (see Fig. 5), aiming at conveying competence on the topic of internet memes while trying to avoid a detrimentally increased level of perceived social distance (status difference) [45] by our subjects. To reduce the animation complexity as well as social cues, the character model[3] wears an oronasal mask avoiding the necessity of lip-synced speech and detailed facial gestures. As masks are common in the COVID pandemic, we do not expect a negative influence by our chosen design. Furthermore, the study was placed in a COVID-sensitive scenario (Sec. 5), e.g., Kate pays attention to COVID-compliant proxemics, thus the mask fits in well. Full-body animations for Kate were taken from Unreal Engine's MCO Mocap Basics[4] and the text-to-speech engine IBM Watson[5] with "Allison V3 with enhanced dnn" (American dialect, default speed, and pitch) was used to generate the various utterances required. To simplify reading, we refer to the ECA as Kate or by the pronoun "she" in the remainder.

4 BEHAVIORAL DESIGN

Following the Media Equation Theory [42], we opted for a plausible, human-like behavior of our ECA. Triggering of utterances and establishing formations are additionally grounded on the environmental requirements given in Section 3.2. If not stated otherwise, the described behavior is true for both conditions (C_{guide} and $C_{companion}$).

As locomotion behavior is our main interest, we identified four logical units for visiting a museum to structure the following descriptions: ECA waiting for a guest (Sec. 4.1), ECA walking with the guest to an exhibit (Sec. 4.2), both standing at an exhibit (Sec. 4.3), and, finally, ECA saying goodbye (Sec. 4.4). To this end, our algorithmic descriptions focus on the ECA's locomotion-, gaze-, and formation-related behavior. However, as speech is not only essential to inform and to express personality, but also to explain or emphasize certain locomotion behavior, we also state utterances of our study for sake of completeness. To convey personality, we, e.g., added personal (e.g., "I especially like this one"), humorous (e.g., "Please follow me, science depends on it.") or relateable (e.g., "It seems cuteness never loses its appeal.") statements for Kate, while locomotion behavior is explained by statements such as "I have to ask you to follow me".

To adjust the ECA's behavior, our algorithms provide various parameters. Informal testing with colleagues in our study setup allowed us to optimize the ECA's behavior for our use case.

[3]The model represents Kate Marsh from "Life is Strange" by Square Enix, available here: https://devhub.vr.rwth-aachen.de/VR-Group/Kate_FBX
[4]https://www.unrealengine.com/marketplace/en-US/product/28fc3cc4332541e3b0037d67a65e5d6d
[5]https://www.ibm.com/cloud/watson-text-to-speech

4.1 First Encounter

In contrast to the first encounter between a user and an ECA in [38], our ECA shares the user's intention to initiate a conversation. However, to not hassle the user, he or she still initiates the first encounter, based on proxemics, i.e., the interpersonal distance between both interactants given as Euclidean distance. To this end, we introduce an invisible, circular awareness zone [10] around the ECA with a radius of $5m$, reflecting the public zone of Hall [23]. On entering this zone, the user triggers the conversation preparation. The ECA engages in mutual gaze and starts to approach the user. Although the average walking speed of adults is $1.4m/s$ [5], informal testing resulted in $1.8m/s$. This slight speed up seemed more natural. The ECA stops at a COVID-compliant, interpersonal distance of $2m$, reflecting the personal space [23], which automatically triggers the ECA's introduction. Identical information across both conditions were the ECA's name, her position as curator as well as the information given about the museum and the current exhibition. The ECA's role description, of course, differed, once being "your tour guide today" (C_{guide}), once stating that she "will accompany you as you will explore our exhibition" ($C_{companion}$).

4.2 Dynamic Transitions

Start of Exploration. To trigger the transition to the first exhibit, utterances are used. For C_{guide}, the ECA finishes her introduction with "Please follow me as we visit some examples from the archives." and walks towards her first goal defined in the annotated environment with approximately $2.0m/s$ (based on internal testing), while footstep sound is embedded to provide locational feedback [7]. In $C_{companion}$, she prompts the user to go ahead with the words "Feel free to look around and I will provide information wherever I can." Then she waits patiently for the user to walk towards the exhibits and follows.

Personal Space (PS). While walking together in either condition, the ECA respects the user's PS. For this dynamic situation, we slightly shrank the interpersonal distance compared to the static first encounter by $0.5m$ to, still COVID-compliant, $1.5m$, based on findings in [8]. Furthermore, we simplified the elliptical shape of the PS found by [8] to a circular safety zone. This is based on the fact that the ellipse found was not very distinct. In addition, subjects in [8] were confronted with a stranger, while the user and the ECA in our scenario are supposed to establish a relationship, also allowing smaller interpersonal distances while still feeling comfortable.

C_{guide}. The ECA guides the user in a predefined order, encoded in the annotated environment, through the museum. To this end, she

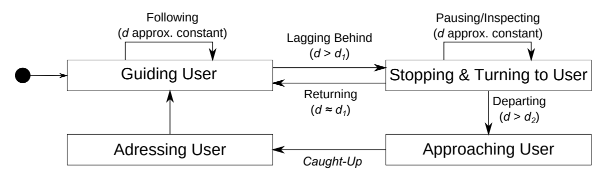

Figure 2: Interplay of ECA (boxes) and user (arrow labels) in the dynamic transitions of C_{guide} with d denoting the interpersonal distance between both interactants, while d_i are certain thresholds to be defined per use case.

(a)

(b)

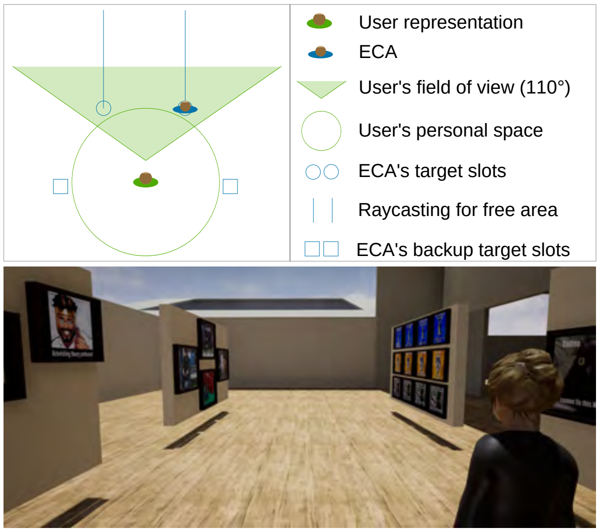

Figure 3: Dynamic transitions in $C_{companion}$ with ECA occupying the right frontal target slot: (a) Sketch with all four target slots. (b) Respective application screenshot with 110° FoV.

walks in front of the user and takes over the global wayfinding decisions. As illustrated in Figure 2, a user-aware ECA thereby needs to react appropriately to user actions. Three actions can be differentiated by analyzing the interpersonal distance d between both interactants: If the user *follows* the ECA, indicated by a rather constant d, the ECA can continue guiding. If the user *falls behind*, indicated by an increasing d, two potential user actions may be the cause: The user may have stopped to *inspect* something, e.g., an exhibit, or the user may have *departed* from the route in order approach another location. Thus, after a distance threshold d_1 is exceeded, the ECA stops and turns towards the user, signaling a waiting behavior. If the d now remains rather constant, the inspection cause is assumed and the ECA shortly asks "Are you coming?" and then waits patiently for the user to continue following. If, however, d still increases, the second cause is assumed.

If a threshold d_2 is exceeded, the ECA starts walking towards the user, with an increased speed of up to $3.0m/s$. Thus, she is able to catch up, even if the user navigates with the maximal speed of $2.5m/s$. After approaching the user to a threshold d_3, the ECA addresses the user and asks to follow her again ("I have to ask you to follow me through the museum. Science depends on it."), before resuming the tour.

All three distance thresholds used in this straight-forward approach strongly depend on the environment given. For our study scenario informal testing resulted in $d_1 = 5m$, $d_2 = 8m$, and $d_3 = 2.5m$.

On reaching the next exhibit, the ECA stops at a predefined location, waits for the user, and then starts her explanations (Sec. 4.3).

$C_{companion}$. The ECA accompanies the user who is solely responsible for the global wayfinding decisions. On detecting the user's interest in a certain exhibit, the ECA will provide information as done in C_{guide}. Since the ECA can be considered as temporarily required assistant [9], we opted for an omnipresent ECA with a short approaching time [7], defined here as time span between recognizing the user's interest in an exhibition and her start of the respective explanation. In contrast to [7], the ECA in our scenario is a companion with whom users should establish a relationship. This *With* [20], defined as group of at least two members maintaining ecological proximity, perceived as being "together", should be reflected in the walking formation of both interactants. A common formation for two people is abreast, defined as moving shoulder-to-shoulder [43]. In VR, especially with HMDs, this spatial layout is unfavorable due to the user's reduced field of view (FoV). Suitable walking positions for the ECA have to meet four requirements: The ECA must (1) be in sight, (2) respect the user's PS, while (3) being close and (4) not occluding the user's direct field of vision. To this end, we decided to define two *target slots* (blue circles, Fig. 3(a)) located towards the peripheral edges of the user's FoV, so adapting the abreast formation to a slight V-shape [43]. Informal testing resulted in slots at $1.5m$ to the front and $0.8m$ to each side added to the user's position.

If the user starts walking, the ECA reacts and aims for the closest slot, trying to stay as close as possible (see Fig. 3(b)). To avoid collisions between the ECA and the scene geometry, a raycast per slot evaluates the presence of obstacles within the next $2m$ (blue lines, Fig. 3(a)). If a slot is occupied by a soon to be collision, the ECA aims for the respective other slot. If there is enough reaction time to change the slots, the ECA passes the user behind his/her back to avoid running in front of his/her feet. If there is not enough time, crossing the user's field of vision is allowed. In case both frontal slots are occupied soon by scene geometry, two backup slots (blue boxes, Fig. 3(a)) at the user's sides are temporarily available, kept outside the user's FoV.

On motion initialization there is a slight delay between the user's and the ECA's start. In addition, if the user turns on the spot, e.g., for 90°, the ECA has to overcome a certain distance to relocate herself near a slot. To, thus, allow the ECA to catch up, her walking speed is adapted with respect to the distance to the closest slot. An exponential relation between distance and speed thereby allows the ECA to quickly catch up while not moving too fast when being close to the user. Informal testing resulted in the speed modulation $v = v_u + a^{(d-c)}$, with the user speed v_u, $a = 1.1$, d defined as distance between ECA and closest slot, and $c = 50cm$. The ECA's speed v is additionally limited to twice the user's maximal speed.

If the user stops, the ECA stops as well and turns towards him or her, signaling a waiting behavior. Facing the user is thereby triggered after a random time in *ms* to show some behavioral variations.

To detect the user's interest in an exhibit, the Euclidean Distance between user and exhibit as well as the user's gaze direction is evaluated. Based on [18], humans tend to keep objects of interest in the center of their field of vision. Thus, we approximate the user's vision

Figure 4: Detecting the user's (represented as mannequin) interest in an exhibit by means of a conical volume of weighted rays from $\omega = 0.2$ (blue) to $\omega = 1$ (red) in steps of 0.2.

Figure 5: ECA Kate explaining memes while altering her gaze between the user (left) and the exhibit (right).

cone by a conical volume of rays moving with the head as shown in Figure 4. The exhibit hit by most rays is considered to be the object of interest. Although a higher density of rays is given for the center of the field of vision due to the conical volume, testing indicated that the result was not accurate enough. Thus, we introduced an additional weighting for the hit count, using a weight of 0.2 for the outer rays up to a weight of 1 for the centric rays. If an exhibit is detected as being in focus for a certain amount of time t and if a certain distance threshold d is not exceeded, the ECA positions herself at a predefined location and starts her explanation, as detailed in the following section. Informal testing resulted in $t = 1s$ and $d = 4m$.

4.3 At an Exhibit: Formation & Interruptions

While the user is free to choose any spot in front of the exhibit, we limit the ECA to a set of locations defined in the annotated environment. We used up to six locations, mostly located on a semi-circle. In C_{guide} the ECA walks towards one of the predefined lateral locations, allowing the user a centric position. In $C_{companion}$, the ECA chooses one of the locations such that one possible location remains free between her and the user. By this, the ECA adds a safety distance to avoid relocating herself in case the user still moves a bit.

Once both interactants are gathered in front of the exhibit, the ECA starts the respective explanatory utterance. If it is the first exhibit of the current room, she automatically prepends the utterance introducing the theme of the respective room as preamble. While explaining the exhibit, Kate alternates her gaze between the user and the exhibit (see Fig. 5). For the exhibit she chooses the *focal point* closest to the user. By this, she is always turned slightly towards the user, inviting him or her to listen. By alternating her gaze, Kate furthermore retains the user's attention to the current exhibit [41].

On finishing her utterance, she pauses a moment in C_{Guide} to allow the user a last look before telling him or her that they are going on. Then she triggers the transition to the next exhibit. In $C_{Companion}$, the ECA waits patiently for the user to resume the free exploration.

While explaining the exhibit, the ECA remains user-aware and is able to react to three types of user actions. First, in case the user's attention wanders, recognized by the conical volume approach described above, the ECA pauses her utterance, addresses the user directly, by "Excuse me, am I boring you?" in C_{guide} and "Actually I wasn't quite finished" in $C_{companion}$, before resuming her explanations. Second, in case of an approaching user who undercuts a distance threshold of $1.2m$, she repositions herself to maintain her PS and stick to a COVID-compliant distance. Thereby, she pauses her utterance and resumes with the last word after relocating. For the relocation, she either moves to the next predefined location away from the user, or she locates herself on a central location if she is already located at the last spot. For the latter, she passes the user's back, while respecting the user's PS. A detour around the exhibit is only taken if the user blocks all other possible trajectories to the new location. Third, in case the user moves on and the distance towards the ECA increases, her reaction depends on the behavioral condition chosen: If the interpersonal distance exceeds a distance threshold d_1 for C_{guide}, the ECA pauses her explanatory utterance. She tries to reengage the user by informing him or her, that she is still explaining, by "Excuse me, am I boring you?", before resuming her explanation. If the user stops and returns to the exhibit, the utterance continues as planned. In case the user departs further exceeding a threshold d_2, she will catch up, ask him or her to follow her again before returning to the exhibit to resume her explanation. This behavior is equal to the catch-up used in the dynamic transitions. Both distance thresholds strongly depend on the environment given. For our study, informal testing resulted in $d_1 = 5m$ and $d_2 = 8m$ due to the sparse scene layout.

Exceeding d_1 for $C_{companion}$ will trigger the same behavior as in C_{guide}, but with "Actually I wasn't quite finished.". However, when d_2 is exceeded, the ECA abandons the exhibit and follows the user with the utterance "Ok, let us look at something else then.".

4.4 Farewell

For C_{guide}, the ECA automatically walks to the tour's final spot after the last exhibition as defined by the annotated environment. There she turns towards the user, engages in mutual gaze, which triggers her farewell utterance. For our study implementation the ECA states "This concludes our tour. I hope you enjoyed our exhibition and maybe you have even found an interesting piece of knowledge to take home. Goodbye and stay safe." Then she walks off. In contrast to our study, where the VR experience ends at this point, the user typically has now the opportunity to revisit the museum on his or her own.

For $C_{companion}$ no definite end of the tour, in terms of time and location, can be defined in advance, as it strongly depends on the user's actions and intents. Thus, users can quit the free, however, accompanied exploration at any time on their own. Therefore, they have to engage in mutual gaze with the ECA and press a designated button on the input device.

5 EVALUATION

5.1 Subjects

Due to the COVID-19 health and safety restrictions of our lab, our user study had to be conducted in a small within-subjects design to decrease the amount of visitors. Eight subjects (7 males, ages M=28.875, SD=1.899), recruited via a university mailing list, participated in the study. All were right-handed and had normal motor skills. While seven subjects had normal or corrected-to-normal vision, deuteranopia occurred once, which was, however, not detrimental for our study. All subjects stated that they had already experienced virtual agents in IVEs, while six subjects also experienced one or two real-life guided tours. All subjects were naïve to the purpose of the study.

5.2 Procedure

On arrival, subjects were informed about the procedure of the study, gave their informed consent and filled out the demographics questionnaire. After the experimenter introduced them to the health and

Table 1: Subjects' ratings wrt the behavioral conditions (M denotes the mean, SD the standard deviation, Mdn the median)

Shortened Statements	C_{guide}			$C_{companion}$		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*
Kate was easily understandable	1.625	0.484	2	1.875	0.331	2
Kate appeared competent	1.75	0.433	2	1.75	0.433	1.5
Kate was entertaining	0.0	1.225	0.5	0.625	0.857	1
Kate caused discomfort	-1.25	1.09	-2	-1 .0	1.5	-2
Kate caused unease	-0.5	1.581	-1	0.375	1.317	0.5
Kate respected PS	1.50	0.707	2	0.75	1.392	-1
Kate positioned logically at exhibit	1.0	1.322	1.5	0.5	1.0	0
Kate behaved weirdly at times	-1.625	0.696	-2	0.5	1.323	0.5
Kate reacted generally appropriately	0.875	0.781	1	1.375	0.484	1
Tour was well paced	0.876	0.927	1	*not applicable*		
Enjoyed being guided	0.625	1.218	1			
Enjoyed free exploration	*not applicable*			1.5	1	2
Walking with Kate felt natural				-0.875	1.053	-1
Kate determined interest correctly				1.25	0.662	1

safety regulations, they were immersed in the training room (red area in Fig. 1). As Kate's mask automatically functions as introduction to a COVID-sensitive scenario, we decided to mimic the real safety precautions customary at the time of our study in VR. To this end, Kate spoke to them over a virtual radio, informing them that the museum was currently cleaned and that they should use the short waiting time to pick up items required for the exploration time, e.g., hand sanitizer, an oronasal mask, and a brochure. By this cover story, subjects got familiar with the ECA's voice as well as with the controller-based navigation as they had to collect the items spread over the room. After picking up all items asked for, doors opened (marked blue in Fig. 1(a)), allowing a transition from the training room into the entrance corridor of the museum. Here, Kate was waiting for the subjects (location 1 in Fig 1(a)) and after approaching her, the study began. The order of C_{guide} and $C_{companion}$ was randomized and counterbalanced. Per condition, Kate introduced herself and the museum briefly (Sec. 4.1), followed by the respective exploration of the museum. After each condition, subjects were asked to take off the HMD to rate their experience by means of a questionnaire containing the Slater-Usoh-Steed (SUS) presence questionnaire [48], the Social Presence Survey (SPS) [2] as well as some questions regarding preferences for the ECA's behavior. Before the second condition had to be experienced, subjects were allowed to do a short break from up to five minutes before being immersed again. This time, they started directly in the entrance hallway while being asked to approach Kate. After experiencing C_{guide} and $C_{companion}$, subjects were asked to fill out a post-study questionnaire with additional free-text fields focusing on comparing both conditions. Afterward, they were offered some chocolate as a reward and left. In total the study took about 60 min/subject, from which 36 minutes were spent fully immersed on average. C_{guide} took thereby about 20min.

5.3 Results

For the standardized questionnaires, we used the proposed 7-point scales (SUS: 1 to 7; SPS: −3=*strongly disagree* to 3=*strongly agree*). For our complementing questions, we used a 5-point scale (−2=*strongly disagree* to 2=*strongly agree* when considering one condition only, 1= C_{guide} to 5=$C_{companion}$ when comparing both).

The mean SUS score for the reported sense of feeling present in the IVE was *M*=4.30, *SD*=0.943 for C_{guide} and *M*=4.10, *SD*=0.985 for $C_{companion}$, indicating a reasonably high level of presence [48].

Table 2: Preference ratings of our eight subjects.

Shortened Statements	Answer Freq.	*M*	*SD*	*Mdn*
… considered superior for learning.		2.75	1.561	2
I felt more comfortable with …		2.5	1.414	2.5
… was more enjoyable.		3.25	1.561	3

C_{guide} 1 2 3 4 5 $C_{companion}$

For computing the SPS score, we used the slight modification proposed in [7]. An average SPS score of *M*=-1.125, *SD*=6.489 for C_{guide} and *M*=-2.125, *SD*=7.524 for $C_{companion}$ was reported.

Subjects were asked to rate their experience based on various questions directly after experiencing a condition. Table 1 summarizes representative results, which were not statistically evaluated due to the low number of subjects. No effects of order were found in the data.

The comparative questions asked after experiencing both conditions, demonstrate two camps of opinion as subjects either voted for the guide or the companion, shown in Table 2. Twice, we found slight tendencies towards the guide and once towards the companion.

Finally, Figure 6 shows the trajectories taken by the ECA and the subjects pooled over all runs from the start to the last exhibition room. As expected due to the predefined order of exhibits, the trajectories are more densely packed for C_{guide} compared to $C_{companion}$.

6 DISCUSSION, IMPLICATIONS & NEXT STEPS

Our goal was to find indicators whether subjects prefer exploratory freedom ($C_{companion}$) over being guided (C_{guide}). Each condition was, however, preferred by four subjects, independent of the presentation order, not supporting **H1**. We also found a large *SD* for both SPS scores, strengthening that the subjects were inconclusive about how convincing the ECA was as a human being. Nevertheless, **H2** can be confirmed, indicating that both behaviors influenced the subjects' perceived (social) presence equally. Besides, the results provide valuable insight on benefits and disadvantages of both roles in the context of exploring social and cultural learning environments.

$\boldsymbol{C_{guide}}$ was rated overall positive. Kate was well understood and was perceived as competent, respected the subjects' PS and didn't cause discomfort or strong unease. Even her rude statement "Excuse me, am I boring you?" was rated neutral in terms of appropriateness, while the five subjects who heard it agreed that Kate had a reason to reprimand them. The majority of subjects enjoyed being guided by her, although the entertainment level was rated rather neutral, while

Figure 6: ECA's and subjects' trajectories for all runs visualized as heat map (normalized to $[0, 1]$, linear color mapping gradient from transparent (0) over red ($1/3$) and yellow ($2/3$) to white (1)); Rotated $90°$ counter-clockwise compared to Fig. 1.

Figure 7: Examples of an ECA's (blue) and a subject's (green) trajectories through certain exhibit areas.

the tour pace seemed to be appropriate. Interestingly, however, the mean SPS score is surprisingly low, characterizing the ECA rather robotic than human-like. This can be explained by some statements of the subjects: Although some stated that they liked the structured experience, the compulsion to focus on one particular exhibit at a time, and that they were confident to not miss crucial content, they also mentioned the very limited interactivity and the monotone, however, robust, character of the tour. Approaching Kate too closely resulted in pauses, which felt unnatural for some subjects, while others stated that they missed being able to skip content as possible with a real human or interactive media. One subject even criticized that the paths taken by Kate were always straightforward. Comparing the ECA's trajectories with the routes taken by the subjects affirms this observation. The ECA's paths are optimized for shortest routes, resulting in walking close to walls or straight to another exhibit, while subjects show more variations and deviations in their routes.

The ECA in $\boldsymbol{C_{companion}}$ was also rated positive. However, some scores were slightly lower. We assume this is caused by her locomotion behavior. Although subjects enjoyed the free exploration and that the ECA was able to determine their interest correctly, walking with Kate felt rather unnatural, resulting in an even lower SPS score compared to C_{guide}. Some subjects stated that it felt like pushing the ECA around. Others were confused when the ECA "suddenly" popped up in their peripheral vision when swapping sides. Despite the embedded footstep sound, these subjects assumed that the ECA was teleporting when being out of sight. Some subjects even experienced inconvenient behavior when passing through rather narrow corridors, as shown in Figure 7(a). Instead of following behind, the ECA ran around the circular arrangement of panels and met the subject unexpectedly at the other side. Despite this unnatural behavior, some subjects experienced $C_{companion}$ as more entertaining compared to C_{guide}. They also enjoyed being in control by exploring the museum in their own pace, taking their time at interesting exhibits and skipping uninteresting ones. The interactivity and the non-linear structure were highlighted several times, while one subject mentioned to be unsure whether he or she missed something. Comparing the subjects' heat maps in Figure 6, however, reveals, that a comparable amount of time was spent at the important locations. In addition, some rooms were even explored more (cp. Fig. 6, room 5 at the top) compared to C_{guide}. It is also noticeable, that subjects used diverse visiting patters (cp. Fig. 7(b), (c)) when exploring the scene. Interestingly, seven subjects were reprimanded with "Actually I wasn't quite finished.", compared to five in C_{guide}. Again, they stated that it was justified, however, rated the appropriateness slightly lower. A first assumption is that subjects considered their inattentiveness as impolite thus accepting the reprimand, while being more sensitive to events limiting their freedom of actions granted by the free exploration.

Based on the results, $C_{companion}$ needs an **improvement** by optimizing walking as a *With* [20]. More group formations in addition to abreast [43] are required while developing strategies to reintroduce the ECA when being temporarily out of sight. Considering visiting styles of users [14] might support anticipating the route to be taken, allowing an extrapolation in addition to a reactive local wayfinding.

Although the design and sample of the study are not optimal, the results yield the following two **implications**: (1) The carefully designed museum combined with our ECA are a feasible system to conduct behavioral studies. (2) Neither a guide nor a knowledgeable companion meet all requirements users have towards a pedagogical agent in a social and cultural application. Therefore, a compromise between both conditions might be best suited for a larger user acceptance in VR, allowing an interactive and adaptable, however, structured learning experience. Users should be able to influence the structure of their experience to a certain degree, while the ECA ensures that no critical information is missed. In case of the museum, users may be guided from room to room by the ECA to deal with the topics in a structured way, while exploring the rooms themselves in an accompanied fashion. Leaving the room may then be negatively commented by the ECA if not all exhibits were seen. Furthermore, the level of detail for the explanations should be controllable, either directly by the user via natural language or by an intelligent ECA interpreting the user's behavior, e.g., his or her attention level, at an exhibit either as bored or interested. A short summary should then be available for uninterested users, while a detailed explanation should be given to curious ones.

There are several avenues for **future work**: After designing an improved companion, a follow-up study with a larger sample size in a between-subject design will give further insights. This second study can also be used to further evaluate the ECA's role on spatial and object memory as well as knowledge transfer, important aspects which have been consciously neglected in this initial work. It may also allow to investigate gender biases, while the ECA's gender and the museum's exhibition theme need to be taken into account during the interpretation of the results. Another important step is to design and evaluate the new ECA's role, detailed above, and to evolve the idea further. One aspect for extending the setting of future studies is adding more social touch: Museums and other scenarios in the context of social and cultural applications deal with communal places, thus enlivening those places adds plausibility and realism to the overall simulation. To this end, virtual staff and visitors should be included to analyze the impact of distractors and external influences on the user's experience as well as to meet the user's nature of being a social being.

7 CONCLUSION

Using ECAs as pedagogical agents in VR-based learning environments becomes more frequent. Focusing on the use case of a virtual museum, we thus evaluated user preferences regarding the ECA's concrete role: While C_{guide} realizes the ECA as virtual guide, limiting the user's role to a passive listener, $C_{companion}$ realizes the ECA as knowledgeable companion, allowing the user a free, accompanied exploration. The results of a within-subjects pilot study yielded no clear preferences to either role, while a promising idea for combining both roles can be derived. The next step will thus be designing an

ECA for an interactive and adaptable, however, structured learning experience and evaluating the ECA's behavior in terms of comfort, knowledge transfer, and spatial knowledge.

REFERENCES

[1] E. Attree, B. Brooks, F. Rose, T. Andrews, A. Leadbetter, and B. Clifford. 1996. Memory Processes and Virtual Environments: I Can't Remember What Was There, But I Can Remember How I Got There. Implications for People with Disabilities. In *Europ. Conf. Disability, Virtual Real. & Associated Techn.*

[2] J.N. Bailenson, J. Blascovich, A.C. Beall, and J.M. Loomis. 2003. Interpersonal Distance in Immersive Virtual Environments. *Personal. Soc. Psychol. Bull.* (2003). https://doi.org/10.1177/0146167203029007002

[3] Joseph Bates. 1994. The Role of Emotion in Believable Agents. *Commun. ACM* 37, 7 (July 1994), 122–125. https://doi.org/10.1145/176789.176803

[4] A. Bogdanovych and T. Trescak. 2017. To Plan or Not to Plan: Lessons Learned from Building Large Scale Social Simulations. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/978-3-319-67401-8_6

[5] R.W. Bohannon. 1997. Comfortable and Maximum Walking Speed of Adults Aged 20-79 Years: Reference Values and Determinants. *Age & Ageing* 26, 1 (1997), 15–19.

[6] A. Bönsch, A. Bluhm, J. Ehret, and T. Kuhlen. 2020. Inferring a User's Intent on Joining or Passing by Social Groups. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1145/3383652.3423862

[7] A. Bönsch, J. Hoffmann, J. Wendt, and T. Kuhlen. 2019. Evaluation of Omnipresent Virtual Agents Embedded as Temporarily Required Assistants in Immersive Environments. In *IEEE Virtual Humans and Crowds for Immersive Environments.* https://doi.org/10.1109/VHCIE.2019.8714726

[8] A. Bönsch, S. Radke, J. Ehret, U. Habel, and T. Kuhlen. 2020. The Impact of a Virtual Agent's Non-Verbal Emotional Expression on a User's Personal Space Preferences. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1145/3383652.3423888

[9] A. Bönsch, T. Vierjahn, and T. Kuhlen. 2017. Evaluation of Approaching-Strategies of Temporarily Required Virtual Assistants in Immersive Environments. In *IEEE Symp. 3D User Interfaces.* https://doi.org/10.1109/3DUI.2017.7893319

[10] A. Bönsch, B. Weyers, J. Wendt, S. Freitag, and T. Kuhlen. 2016. Collision Avoidance in the Presence of a Virtual Agent in Small-Scale Virtual Environments. In *IEEE Symp. 3D User Interfaces.* https://doi.org/10.1109/3DUI.2016.7460045

[11] S. Cairncross and M. Mannion. 2001. Interactive Multimedia and Learning: Realizing the Benefits. *Innov. in Edu. and Teaching Intern.* (2001). https://doi.org/10.1080/14703290110035428

[12] L. Cao, J. Lin, and N. Li. 2019. A Virtual Reality Based Study of Indoor Fire Evacuation After Active or Passive Spatial Exploration. *Comput. Human Behav.* (2019). https://doi.org/10.1016/j.chb.2018.08.041

[13] M. Carrozzino, M. Colombo, F. Tecchia, C. Evangelista, and M. Bergamasco. 2018. Comparing Different Storytelling Approaches for Virtual Guides in Digital Immersive Museums. In *Intern. Conf. Augmented Real., Virtual Real., and Comput. Graphics.* https://doi.org/10.1007/978-3-319-95282-6_22

[14] L. Chittaro and L. Ieronutti. 2004. A Visual Tool for Tracing Users' Behavior in Virtual Environments. In *Proc. Work. Conf. Adv. Vis. interfaces.* https://doi.org/10.1145/989863.989868

[15] E.R. Chrastil and W.H. Warren. 2013. Active and Passive Spatial Learning in Human Navigation: Acquisition of Survey Knowledge. *J. Experimental Psych.: Learning, Memory, and Cognition* (2013). https://doi.org/10.1037/a0032382

[16] I. de Kok, J. Hough, F. Hülsmann, M. Botsch, D. Schlangen, and S. Kopp. 2015. A Multimodal System for Real-Time Action Instruction in Motor Skill Learning. *Proc. ACM Int. Conf. Multimodal Interact.* (2015). https://doi.org/10.1145/2818346.2820746

[17] P. Doyle and B. Hayes-Roth. 1997. Guided Exploration of Virtual Worlds. In *Network and Netplay: Virtual Groups on the Internet.* 243–264.

[18] T. Foulsham, E. Walker, and A. Kingstone. 2011. The Where, What and When of Gaze Allocation in the Lab and the Natural Environment. *Vision Research* (2011). https://doi.org/10.1016/j.visres.2011.07.002

[19] N. Fourati, A. Richard, S. Caillou, N. Sabouret, J.C. Martin, E. Chanoni, and C. Clavel. 2016. Facial Expressions of Appraisals Displayed by a Virtual Storyteller for Children. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/978-3-319-47665-0_21

[20] E. Goffman. 2009. *Relations in Public.* Transaction Publishers.

[21] C. Goodwin. 1980. Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Sociological Inquiry* (1980). https://doi.org/10.1111/j.1475-682X.1980.tb00023.x

[22] J. Gratch, D. DeVault, and G.M. Lucas. 2016. The Benefits of Virtual Humans for Teaching Negotiation. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/978-3-319-47665-0_25

[23] E.T. Hall. 1968. Proxemics. *Current Anthropology* 9, 2-3 (1968), 83–96.

[24] T. van der Heiden, N.N. Shankar, and H.V. Hoof. 2020. *Social Navigation with Human Empowerment Driven Reinforcement Learning.* Technical Report.

[25] L. Hoffmann, N.C. Krämer, A. Lam-Chi, and S. Kopp. 2009. Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent?. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/978-3-642-04380-2_19

[26] J. Ibanez, R. Aylett, and R. Ruiz-Rodarte. 2003. Storytelling in Virtual Environments from a Virtual Guide Perspective. *Virtual Real.* (2003). https://doi.org/10.1007/s10055-003-0112-y

[27] J. Ibanez, R. Aylett, and R. Ruiz-Rodarte. 2003. That's my Point! Telling Stories from a Virtual Guide Perspective. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/978-3-540-39396-2_42

[28] K. Isbister and P. Doyle. 1999. Touring Machines: Guide Agents for Sharing Stories about Digital Places. In *Proc. Workshop on Narrative and Artificial Intell.*

[29] Ryo Ishii and Yukiko I Nakano. 2008. Estimating User's Conversational Engagement based on Gaze Behaviors. In *International Workshop on Intelligent Virtual Agents.* Springer, 200–207.

[30] D. Jan, A. Roque, A. Leuski, J. Morie, and D. Traum. 2009. A Virtual Tour Guide for Virtual Worlds. *Proc. Int. Conf. Intell. Virtual Agents)* (2009). https://doi.org/10.1007/978-3-642-04380-2_40

[31] D. Karreman, G. Ludden, and V. Evers. 2015. Visiting Cultural Heritage with a Tour Guide Robot: A User Evaluation Study in-the-Wild. In *Social Robotics.* https://doi.org/10.1007/978-3-319-25554-5_32

[32] D. Karreman, E. van Dijk, and V. Evers. 2012. Using the Visitor Experiences for Mapping the Possibilities of Implementing a Robotic Guide in Outdoor Sites. In *IEEE Intern. Symp. Robot and Human Interactive Commun.* https://doi.org/10.1109/ROMAN.2012.6343889

[33] S. Kopp, L. Gesellensetter, N. Krämer, and I. Wachsmuth. 2005. A Conversational Agent as Museum Guide–Design and Evaluation of a Real-World Application. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/11550617_28

[34] T. Kruse, A. Pandey, R. Alami, and A. Kirsch. 2013. Human-aware Robot Navigation: A Survey. *Rob. Auton. Syst.* (2013). https://doi.org/10.1016/j.robot.2013.05.007

[35] M. Lim and R. Aylett. 2007. Feel the Difference: A Guide with Attitude!. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/978-3-540-74997-4_29

[36] S. Liszio and M. Masuch. 2016. Lost in Open Worlds: Design Patterns for Player Navigation in Virtual Reality Games. In *Proc. Intern. Conf. Advances in Comput. Entertainment Techn.* https://doi.org/10.1145/3001773.3001794

[37] M. Numfu, A. Riel, and F. Noël. 2019. Virtual Reality based Digital Chain for Maintenance Training. *Procedia CIRP* (2019). https://doi.org/10.1016/j.procir.2019.04.268

[38] S. Ólafsson, B. Bédi, H. Helgadóttir, B. Arnbjönsdóttir, and H. Vilhjálmsson. 2016. Starting a Conversation with Strangers in Virtual Reykjavik: Explicit Announcement of Presence. In *Proc. from 3rd Eur. Symp. Multimodal Commun.* 62–68.

[39] Z. Pan, W. Chen, M. Zhang, J. Liu, and G. Wu. 2009. Virtual Reality in the Digital Olympic Museum. *IEEE Comp. Graphics and Applications* (2009). https://doi.org/10.1109/MCG.2009.103

[40] A. Pandey and R. Alami. 2009. A Step Towards a Sociable Robot Guide which Monitors and Adapts to the Person's Activities. In *IEEE Intern. Conf. Advanced Robotics.* 1–8.

[41] N. Pfeiffer-Lessmann, T. Pfeiffer, and I. Wachsmuth. 2012. An Operational Model of Joint Attention - Timing of Gaze Patterns in Interactions between Humans and a Virtual Human. In *Proc. Annual Conf. Cognitive Science Society.*

[42] B. Reeves and C. Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.*

[43] F.A. Rojas and H.S. Yang. 2013. Immersive Human-in-the-Loop HMD Evaluation of Dynamic Group Behavior in a Pedestrian Crowd Simulation that Uses Group Agent-based Steering. *Proc. ACM SIGGRAPH Int. Conf. Virtual Real. Contin. Its Appl. Ind.* (2013). https://doi.org/10.1145/2534329.2534336

[44] R. Rzayev, G. Karaman, K. Wolf, N. Henze, and V. Schwind. 2019. The Effect of Presence and Appearance of Guides in Virtual Reality Exhibitions. In *Proc. Mensch und Computer.* https://doi.org/10.1145/3340764.3340802

[45] S. Sylaiou, V. Kasapakis, E. Dzardanova, and D. Gavalas. 2019. Assessment of Virtual Guides' Credibility in Virtual Museum Environments. In *Augmented Real., Virtual Real., and Comput. Graphics.* https://doi.org/10.1007/978-3-030-25999-0_20

[46] D. Traum, P. Aggarwal, R. Artstein, S. Foutz, J. Gerten, A. Katsamanis, A. Leuski, D. Noren, and W. Swartout. 2012. Ada and Grace: Direct Interaction with Museum Visitors. In *Proc. Int. Conf. Intell. Virtual Agents.* https://doi.org/10.1007/978-3-642-33197-8_25

[47] I. Tussyadiah, D. Wang, T. Jung, and M. Dieck. 2018. Virtual Reality, Presence, and Attitude Change: Empirical Evidence from Tourism. *Tourism Management* (2018). https://doi.org/10.1016/j.tourman.2017.12.003

[48] M. Usoh, E. Catena, S. Arman, and M. Slater. 2000. Using Presence Questionnaires in Reality. *Presence: Teleoperators & Virtual Environments* (2000). https://doi.org/10.1162/105474600566989

[49] A. Velentza, D. Heinke, and J. Wyatt. 2019. Human Interaction and Improving Knowledge through Collaborative Tour Guide Robots. In *IEEE Intern. Symp. Robot and Human Interactive Commun.* https://doi.org/10.1109/RO-MAN46459.2019.8956372

[50] M. Yousuf, Y. Kobayashi, Y. Kuno, K. Yamazaki, and A. Yamazaki. 2019. Social Interaction with Visitors: Mobile Guide Robots Capable of Offering a Museum Tour. *IEEE Trans. Electrical and Electronic Engrg* (2019). https://doi.org/10.1002/tee.23009