International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2023)
J.-M. Normand, M. Sugimoto, and V. Sundstedt (Editors)

Embodied Conversational Agents with Situation Awareness for Training in Virtual Reality

P. Kán, M. Rumpelnik, and H. Kaufmann

TU Wien, Institute of Visual Computing and Human-Centered Technology, Austria

Figure 1: *Left: User's view of our embodied conversational agent in VR training. Right: User immersed in our VR environment.*

Abstract

Embodied conversational agents have a great potential in virtual reality training applications. This paper investigates the impact of conversational agents on users in a first responder training scenario. We integrated methods for automatic speech recognition and speech synthesis with natural language processing into a VR training application in the Unity game engine. Additionally, we present a method for enabling situation awareness for agents in a virtual environment. Finally, we conducted a between-subject lab experiment with 24 participants which investigated differences between conversational agents and agents with pre-scripted audio. Several metrics were measured in the experiment including presence, subjective task performance, learning outcome, interaction quality, quality of information presentation, perceived realism, co-presence, and training task duration. Our results suggest that users trying our conversational agents condition experienced significantly higher level of co-presence than users with pre-scripted audio. Additionally, significant differences in subjective task performance and training duration were discovered between genders. Based on the results of our qualitative analysis, we provide guidelines that can facilitate future design of VR training applications and research studies with embodied conversational agents.

CCS Concepts

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

1. Introduction

Virtual embodied agents play an important role in virtual reality (VR) training applications [CW19]. These computer-actuated characters have the potential to improve realism, increase presence [SBS19], and provide important communication cues [RHW20]. Moreover, they can provide a trainee with additional information which can be considered useful in resolving a given training problem. However, achieving realistic behavior of agents is inherently a difficult problem due to the high level of conversational capabilities of real people which are often lacking in

DOI: 10.2312/egve.20231310

the case of virtual agents. Conversational capabilities are especially useful in VR training for rescue scenarios due to the frequent communication and interaction with civilians during real operations. Agents in VR training applications typically only use prescribed dialogues or pre-recorded speech and they often do not have conversational capabilities. Another important problem of conversational agents in VR is missing situation awareness. The agents are often unable to capture changes in the VR environment or their states and reliably communicate them with a user. Moreover, the impact of conversational agents on users and their task performance in virtual reality training has not yet been extensively studied. Therefore, an investigation of the influence of conversational agents on VR training is required to gain important insights for future development and research of training applications in VR. Previous research studied various aspects of embodied conversational agents (ECA)s in VR [SBS19, WSR19, KBH*18, KdMN*20], however, more studies are required to learn about their impact on training performance, learning outcome, interaction efficiency, and other aspects.

We hypothesize that the conversational capabilities of agents are beneficial for VR training and we investigated their impact on subjective user-provided metrics and on task duration in a user study. A novel system enabling conversational capabilities of virtual agents utilizing automatic speech recognition (ASR) and natural language processing (NLP) was developed in our research to enable our investigation. Additionally, we propose a novel method for enabling situation awareness and state awareness of each agent. Awareness of the agent about the surrounding environment allows the agent to provide relevant information to a trainee. Our method utilizes a connection from speech services back to the Unity application to obtain agent-relevant information about the actual environment.

Our user study investigated the impact of ECAs on training task duration, subjective task performance, learning outcome, interaction with agents, information presentation, realism, presence, and co-presence in a VR rescue scenario (Figure 1). We compared the following two conditions in a between-group experiment: (1) Agents with full conversational capabilities (i.e. users could naturally speak with agents) and (2) agents who provide information by speaking the prescribed text but cannot answer users' questions. Our goal was to find out what is the important benefit of enabling embodied agents with conversational capabilities in comparison to agents which provide all information at once. Both conditions included situation awareness so the agents always reported actual information. In addition to the differences between our two conditions, we also studied gender-related differences. Finally, based on the results of our qualitative analysis, we define guidelines for future design and research of VR training scenarios with ECAs. The results of our study can be useful for future development and research of VR training applications with conversational agents.

2. Related Work

2.1. Training in Virtual Reality

Virtual reality is a useful tool for training in numerous domains including medical training, first responder training, training for industrial maintenance and assembly, and training in the education domain. Previous research investigated how VR can be utilized to achieve the best training outcomes. Multiple aspects of first responder training in VR were investigated including locomotion [MFS*17], perceived task complexity [SSS98], and system usability [LLGPM21]. Additionally, user behavior in search and rescue tasks for firefighter training was studied by Doroudian et al. [DWW*22]. The authors investigated the effect of using immersive maps in VR on user behaviour during training. Virtual reality was also used for emergency evacuation training [SRD15], disaster response training [NJD19], and training of forest firefighters [LPL22]. Several of these works also utilized embodied agents in VR to improve training performance [JBG*19]. However, previous first responder training systems in VR often used simple agents without conversational capabilities or with prescribed dialogues.

An important aspect of VR training is user perception which is typically investigated by user studies. Peretti et al. [PSSE21] studied a gamification approach for first responder VR training. The requirements for VR first responder training were investigated by Haskins et al. [HZG*20]. Moreover, psychophysiological measurements were used by Paletta et al. [PSR*22] to assess levels of stress during training in real and virtual environments. Their results indicate high levels of cognitive-emotional stress in both real and virtual conditions. The authors did not find significant differences between conditions. The influence of interaction technologies on learning in VR assembly training was investigated by Vélaz et al. [VRAG*14]. While the authors did not find a significant impact of interaction on learning, their results suggest that different interaction technologies have a significant impact on training time.

2.2. Conversational Agents in Mixed Reality

With recent progress in the field of artificial intelligence, conversational agents have tremendous potential in numerous application areas including virtual reality and training. Additionally, an embodiment of these agents into a virtual body can have a positive influence on co-presence and other aspects of VR. Therefore, previous research investigated the utilization of embodied conversational agents (ECA)s in various application domains of VR. ECAs for interpersonal skills training, including sales pitching, negotiation, leadership, interviewing, and communicating with empathy, was proposed by Chetty and White [CW19]. Embodied conversational agents were also used to help with navigation and collaboration in virtual environments [BEGGN98]. A VR system for training verbal de-escalation skills by clinicians, utilizing ECAs, was proposed and studied by Moore et al. [MAB*22]. The authors identified important qualitative factors for VR training including agency, accessibility, visibility, privacy, realistic task, information about completion, motion sickness, and others. Previous research on situation awareness of ECAs in VR is sparse. Ijaz et al. [IBS11] proposed a method for situation awareness of ECAs using two levels of annotation: objects annotation and regulation annotation. However, their method is mostly focused on spatial and directional information about objects and it requires complex set of rules.

Different technological solutions were used in the past to enable the embodiment of AI conversational agents into virtual bodies. Typically, algorithms for automatic speech recognition (ASR), natural language processing (NLP), natural language generation (NLG), and speech synthesis are required to enable the conversational capabilities of ECAs. Additionally, animations play an im-

portant role for realistic virtual humans. A framework for the embodiment of conversational agents in VR and AR was proposed by Hartholt et al. [HFR*19]. Griol et al. [GSMC19] developed social embodied agents with conversational capabilities. Their quantitative evaluation showed a 94% success rate in completed dialogues with correct information provided by the agent. A mobile AR game, utilizing ECAs, for sexual assault bystander intervention training was presented by Schlesener et al. [SLB*23]. The authors used IBM Watson Assistant for conversational AI. Traum and Rickel presented an ECA system for mission rehearsal exercises [TR02].

The impact of ECAs on human perception plays an important role also in mixed reality applications, particularly in training. Therefore, human-oriented studies are necessary to learn how ECAs influence various aspects of training in mixed reality. Reinhardt et al. [RHW20] investigated the user preference between a simple and realistic appearance of ECA. In contrast to the theory of Uncanny Valley [MMK12] the results of Reinhardt et al.'s study indicate that people prefer realistic visualization due to the possibility of additional communication features like eye contact or gaze. The authors also recommend implementing social cues (e.g. turning the agent in the direction of the user and following the user with the agent's eyes). In our research, we also integrated turning animation to direct our agents to the user if they are nearby. The importance of the embodiment of an intelligent agent into a virtual body in AR was studied by Kim et al. [KBH*18]. Their results suggest that the embodiment of an agent can increase the user's confidence about the agent's ability to influence the real world and about real-world awareness of the agent, in comparison to the agent without a virtual body. Additionally, an interesting finding of the authors was that agent's embodiment has a positive effect on the user's confidence that the agent will respect the user's privacy. Another study of the embodiment of virtual agents was conducted by Wang et al. [WSR19]. The authors compared four embodiment conditions including voice-only, non-human, full-size human, and miniature embodied agent. Their results indicate users' preference of miniature embodied agent. Specific aspects of ECAs were also studied in the past, including the agent's locomotion metaphor [TRO*19], the agent's capability of real objects manipulation [SNS19], and effects on collaborative decision-making [KdMN*20].

While previous research demonstrated the high potential of ECAs in VR training, the impact of ECAs on users in first responder VR training has not been sufficiently studied. In our research, we conducted a user study that focuses on open questions about the presence, subjective task performance, learning outcome, interaction quality, information presentation, realism, co-presence, and training task duration with ECAs in first responder VR training. Additionally, we present a novel methodology for enabling situation awareness for each embodied agent in a training environment.

3. Embodied Conversational Agents with Situation Awareness

One of the crucial capabilities of ECAs is to maintain natural speech conversation with a user. We achieve this goal using a combination of methods for automatic speech recognition (ASR), natural language processing (NLP), and text-to-speech (TTS). Additionally, we enhanced our VR agents with turning animations, gestures, lip-sync, and spatial sound rendering. One of our contributions is that we enabled our agents with situation awareness using a local database. The conversational pipeline, used for our virtual agents, is depicted in Figure 2. Initially, when a voice input is detected and the user is in the proximity of an agent, the audio is recorded by our Unity application and it is streamed to the NVIDIA Riva server for speech recognition. NVIDIA Riva ASR service translates the audio to text and sends it back to the Unity application. This text is then sent to the Rasa server for NLP. Rasa analyzes the transcript, extracts relevant data, and determines the user's intent. Depending on the determined intent, it may need further information from the Unity VR application and in this case, it makes an HTTP request to the web service on the Unity side. This way we can achieve situation awareness (details in Section 3.2) of our agents because Unity can provide additional information for the generation of the answer. With this additional information, Rasa creates a response text of an agent and sends it to the Unity application as an answer of the specific agent. Finally, Unity queues this response text for the Riva text-to-speech service. The resulting audio file, received by the Unity VR application, is then played to a user. All processes can run either on a single computer or on a distributed setup with a GPU server and a client VR-ready PC.

Our agents are embodied into humanoid bodies from Rocketbox Avatar Library [GFOP*20]. This library also contains turning animation, idle animation, and listening animation which we used to increase the realism of agents in VR. We set each agent to turn towards a user and to start listening when a user enters the radius of n meters surrounding the agent (We empirically set $n = 3$). When the NLP pipeline finishes and an agent is responding to a user with an audio answer, we use 3D spatial sound rendering to enable a user to perceive the direction of the agent's voice coming from the agent's position. Additionally, we use Oculus Lipsync for Unity to enable the synchronisation of the agent's lips with the voice.

3.1. Natural Language Processing

We handle the NLP of user utterances in a Rasa service. Its main task is to analyze input text, detect intents, extract entities, and finally provide an answer. Intents relate to a task that a user wants an agent to perform and entities are pieces of information that are needed to accomplish this task. For example, if the user input to the model is the question "What happened in the factory?", we want the model to detect it as the intent of *what_happened* with the location entity set to *factory*. Intent and entity must be detected by the NLP service. In our method, the NLP service can additionally connect back to Unity application through HTTP call to obtain scene information and enable situation awareness in agent's answers.

Training Data: Our training dataset for Rasa natural language understanding consists of 492 example sentences, which model 27 different user intents with 7 entities. This model was used in our experiments with first responder training in VR. The average training time of our model with 200 epochs was 7 minutes. The training is only required once and then the NLP pipeline can run interactively.

Actions: After receiving a user utterance, the model predicts which action should be executed. Actions can be simple responses that send a text to a user or custom actions that can run any Python code. We are using simple responses as an answer to e.g. "Hello".

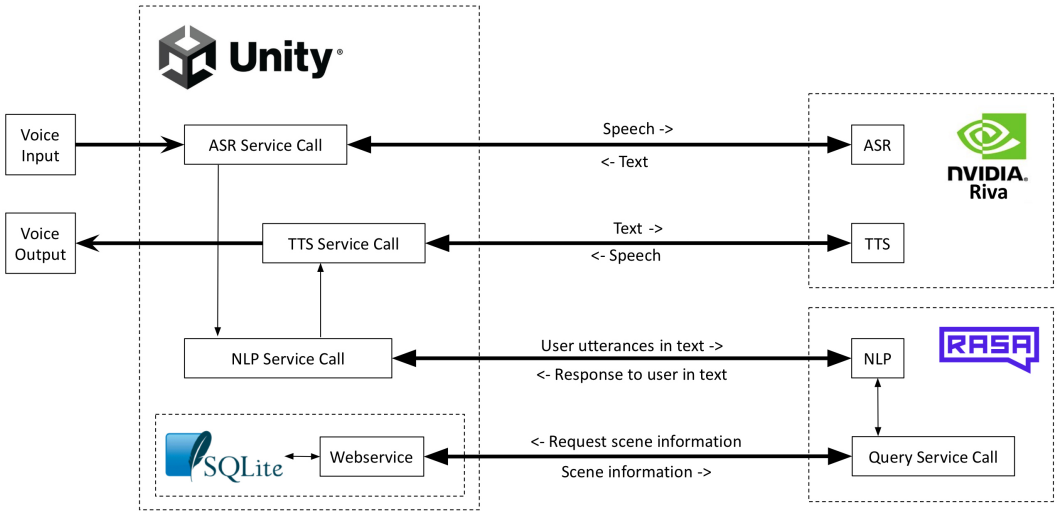

Figure 2: *Schematic representation of our technical solution enabling ECAs in VR, including situation awareness. Situation awareness is enabled by using a back connection from the Rasa service to the Unity VR application which provides the required information about the environment and agents. Services on the right side (NVIDIA Riva and Rasa) can be deployed either locally or in a distributed setup.*

Custom actions are used when scene-specific information is requested such as conditions of agents or incidents at locations. If scene information is required, the Rasa service makes an HTTP request with the parameters of the required information to the web service of our Unity VR application. The web service parses the sent parameters, queries the database, and returns the requested data in JSON format. Each custom action in the Rasa service is then responsible for parsing the JSON data and building a response. For building responses, we have defined 33 template response messages such as "I have *injury_description* at my *bodypart*" or "My name is *firstname lastName*", where the data identifier (with italic font) is replaced by custom data, received from the Unity application.

Stories: Stories are example conversation representations between a user and an agent that are needed for training the dialogue management part of the model. Stories are lists of steps where user inputs are expressed as intents and responses are expressed as actions. Additionally, entities should be listed for given intents because the model learns to predict the next action based on the combination of intent and entities. In the stories, we also need to make sure that we have converging paths dependent on entities that are set. Every story path where a subject entity is extracted during intent recognition needs to determine which specific entity is referred to. For example, let's consider the question: "What happened to you?", the keyword *you* is recognized as a subject and mapped to self. Every agent can answer questions about herself/himself, therefore the agent will not have to ask who is meant by *you*. Now let's consider the question: "What happened to him?": If this is the first question a user asks an agent, the agent has no information about the subject *him* and therefore a story, in which an agent asks a user who is he, must exist. If there was already some conversation between a user and the agent and they already talked about another agent (i.e. the subject *him* is already defined), another story, that uses this information without asking about the name, must exist. We defined 45 different stories with various conversation paths.

3.2. Situation Awareness

We enable situation awareness for our agents by providing information about the agent's surroundings, conditions, and state through the Unity application using an HTTP web service. The VR application acts also as an HTTP server and it can provide information about the scene and agents in JSON format using an HTTP Get call.

The virtual world in our VR training is divided into multiple parts using invisible bounding boxes which represent named areas or locations. Hence the enclosing bounding box for a factory building has the name of the factory. In addition to the location name, we also define a description of where in the scene the location is located so agents can help users to find a location. Just like in the real world, all objects and agents in the scene must be inside a certain location. We use this location division to give agents limited knowledge, so by default, they do not know about the whole scene. By default an agent only knows about her/his enclosing location but we can add additional known locations in the Unity editor. Defining different locations throughout the scene enables questions such as "What happened in the factory?", "Where is the car?" or "How many people are injured inside the hotel?".

The information about the properties, conditions, and states of all agents, incidents, and tools is stored in a local SQLite database. This database is built at the start of the project and it is used throughout the VR session as a centralized source of data about the environment for situation awareness. The initialization of the database is done automatically using provided data from the Unity editor. This way, a programmer can easily edit the properties of agents, incidents, or tools in the training environment. In our model, every agent has the following properties: name, age, gender, and location. Additionally, each agent can have multiple conditions including injuries of different body parts or ability to move. The database also contains information about incidents, their locations, how they can be solved, and which tools should be used to solve

them. The database information is updated as a user proceeds in the training scenario to always reflect the actual state of the environment and agents. This way, the agents are able to communicate about their personal information, their injuries (until healed), and incidents in the environment including their locations and causes. Additionally, agents can provide information on how a specific problem should be solved if intended by a training session. Integration of scene information into natural language responses is done in the Rasa service which invokes an HTTP request to the Unity application if this information is required.

4. User Study

4.1. Study Design

We investigated the impact of embodied conversational agents in a first responder VR training on users in a user study. We were particularly interested in the impact on the sense of presence, subjective task performance, learning outcome, interaction quality, quality of information presentation, perceived realism, co-presence, and training task duration. Our experiment followed a between-group design with two conditions:

1. Conversation: In this condition, the embodied agents in the VR training scenario were capable of speech conversation with a user. The user could obtain the necessary information about the virtual environment by asking questions in natural language. This way a user could discover the problems which needed to be solved, to accomplish the first responder training.

2. No-Conversation: The second condition contained the same virtual world and embodied agents in a first responder VR training scenario as the first condition. However, in this condition, the agents were not able to respond to the questions of the user, but when the user approached them (i.e. entered their interaction radius), the agents revealed all their available information by speech. This was a monologue of an agent towards the user. The training tasks were the same for both conditions (Section 4.2).

All the subjective user metrics were measured using an ad hoc post-experiment questionnaire (Supplementary material). We utilized questions from previous research on presence [WS98, VWG*04] and we also added new items for other metrics. The duration of the training task was measured directly by time measurement from the start of the VR training until all the problems in the environment were resolved (i.e. until all eight tasks were accomplished). Our hypotheses were that Conversation condition achieves better scores in comparison to the No-Conversation condition in the following metrics: (**H1**) Presence, (**H2**) Subjective task performance, (**H3**) Learning outcome, (**H4**) Agents interaction, (**H5**) Information presentation, (**H6**) Realism, (**H7**) Co-presence, and (**H8**) Duration of training task.

4.2. Apparatus

In order to investigate the impact of ECAs on human perception in a first responder VR training, we designed a VR training application in the Unity game engine (Figure 1). The created virtual environment has a size of 62×54 meters with a walkable area of 1688 m^2. The environment contains eight incidents (i.e. eight tasks) that need to be resolved by a trainee in VR. Four incidents are environment-related (chemical waste, fire, locked door, and stuck car door) and the remaining four incidents are four injured agents which need to be healed by the user. The training task of a user is to solve all eight incidents in the VR environment. Our designed VR environment includes tools (first aid kit, tongs, sandbag, fire extinguisher, and axe) that are needed to resolve the incidents and heal the agents. A schematic map of the environment can be seen in supplementary material. The environment is inhabited by eight embodied agents which a user can interact with. Since the agents are aware of their location and surroundings, they can help to navigate users and to discover the incidents. The locomotion of a user in the VR environment is realised using a teleportation metaphor. Users can interact with objects using a touch gesture with a VR controller. When the user's controller is close to an object, the user can press the grip button to grab and hold the object. The user can then use this object to resolve an incident in the environment by touching the incident location with an object. For example, a user can heal injured agents by touching them with the first aid kit. An exception is a fire extinguisher, that interacts with fire by pointing the extinguishing spray toward the fire. A user is informed about the remaining tasks to be solved by a small information panel on her/his wrist. When a user solves all eight tasks, a message is shown in the user's view that informs about the accomplishment of the training. At this time, the VR part of the experiment is finished by the experimenter and the time measurement is taken.

In our pedagogical scenario, a user is exposed to the number of events and incidents that need to be solved. In a real first responder VR training a feedback on user performance (i.e. how well the situation was resolved by the user and which mistakes were made) would be necessary to achieve pedagogical goals. As our research was focused on conversational agents, we did not implement the feedback procedure and we measured the learning outcome by subjective response of the users.

We used a PC with Intel Core i9-11900K CPU and Nvidia RTX 3090 GPU to run the Unity VR application and NVIDIA Riva and Rasa servers. A head-mounted display HTC Vive Pro was used to immerse users in VR. We used headset's integrated microphone and headphones as default audio communication channels.

4.3. Procedure

As we used a between-group design in our study, every participant only experienced one VR training condition. In the beginning, every participant signed informed consent and read an information form about the study. Then a short demographic questionnaire and a simulator sickness questionnaire (SSQ) [KLBL93] were filled out by a participant. Afterward, the user could experience our VR setup in an initial test scenario. This scenario only contained a simple world with few tools and one agent to allow a user to familiarize herself/himself with interaction and navigation metaphors and with communication with an agent (speech conversation was only enabled for the Conversation condition). When a user indicated to be ready, the actual VR training scenario could start. Participants had unlimited time to accomplish all training tasks and the longest duration was 14 minutes and 54 seconds. A user could also finish the training without accomplishing all the tasks but no one used this

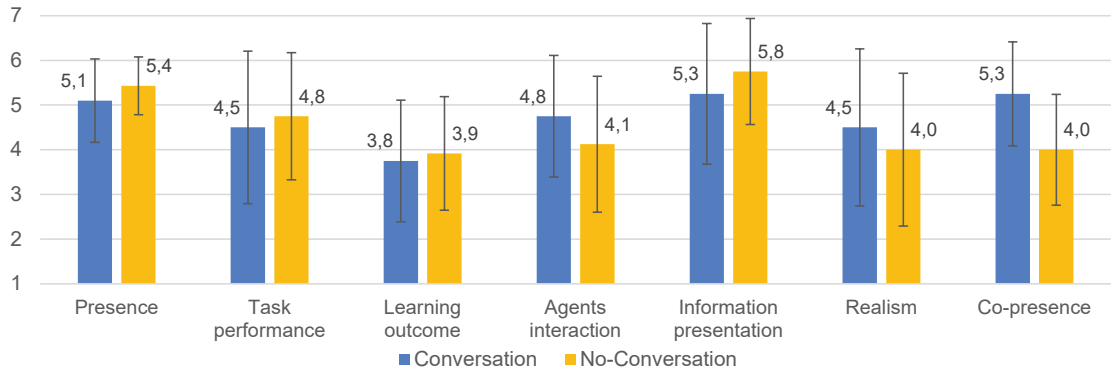

Figure 3: *The results of subjective responses of participants to our questionnaire. The y-axis shows values of the Likert scale and each bar represents the average value of participants' answers for a given condition. A higher value represents a more positive response of participants for a given metric. The error bars indicate standard deviations in responses.*

Table 1: *Significance of differences between two compared conditions, calculated by Mann-Whitney U test. We used the value of .05 for interpreting significance. We can observe significant difference in Co-presence between two studied conditions.*

Metric:	Presence	Task performance	Learning outcome	Agents interaction	Information presentation	Realism	Co-presence	Duration
U value	60.5	68.5	70	55	56.5	60	36	61
p-value	.52	.87	.94	.34	.38	.5	**.035**	.55

option and all participants successfully finished all 8 tasks. After the first responder VR training scenario, each participant filled out again SSQ and additionally our main study questionnaire to measure subjective responses to all designed metrics. The average time, spent in our VR training application, was 7 minutes and 3 seconds.

4.4. Participants

24 participants took part in our user study including 8 females and 16 males. We ensured an equal gender distribution across conditions (i.e. every condition was experienced by 4 females and 8 males). The average age of participants was 32.5 years (SD = 7 years). The majority of participants were from the academic environment including students, researchers, and technicians.

4.5. Results

All participants accomplished the VR first responder training with all eight tasks. The results of participants' answers in our questionnaire for two conditions (1) Conversation and (2) No-Conversation can be seen in Figure 3. We investigated the statistical significance of differences between the two conditions using the Mann-Whitney U test (Table 1). Significant difference between conditions was only found for the Co-presence factor in favor of the Conversation condition. Other measured factors did not elicit statistical significance between conditions. The Conversation condition achieved higher scores in Co-presence, Agents interaction and Realism metrics. For all the other metrics, including Presence, Subjective task performance, Learning outcome, and Information presentation, the No-Conversation condition achieved slightly better scores.

The average VR training completion times (i.e. training task durations) were 7 minutes and 39 seconds (SD = 3 min. and 28 sec.) for the Conversation condition and 6 min. and 28 sec. (SD = 2 min. and 10 sec.) for the No-Conversation condition. As we can see in Table 1, this difference was not statistically significant.

In addition to our main hypotheses, we were interested in differences in measured metrics across genders. The results of the dependency of measured factors on gender can be seen in Figure 4. The respective significance values, calculated by the Mann-Whitney U test, can be seen in Table 2. In this case, subjectively rated task performance was reported significantly higher by male participants than by female participants. Additionally, male participants accomplished VR training significantly faster than female participants (Table 2). The VR training completion times were on average 6 min. and 28 sec. (SD = 3 min. and 6 sec.) for males and 8 min. and 15 sec. (SD = 2 min. and 18 sec.) for females. The other metrics did not elicit statistically significant differences.

Simulator Sickness Questionnaire: We analyzed simulator sickness using a questionnaire (SSQ) from Kennedy et al. [KLBL93]. Each participant filled out SSQ twice (before and after the VR exposure). We used all 16 items, proposed by Kennedy et al. to measure simulator sickness and accumulated them into four factors as suggested by the authors: nausea (N), oculomotor disturbance (O), disorientation (D), and total simulator sickness (TS). The results of measured simulator sickness responses are shown in Table 3. We assessed the statistical significance of the difference in SSQ scores between the pre-test and post-test using the Wilcoxon signed-rank test. None of the differences were significant.

Qualitative Analysis: Our study questionnaire contained three open questions about the training scenario, about the agents and about possible improvements. We analyzed these three questions by extracting the codes for both positive and negative user judgements. The results of this open coding can be seen in Table 4. As we

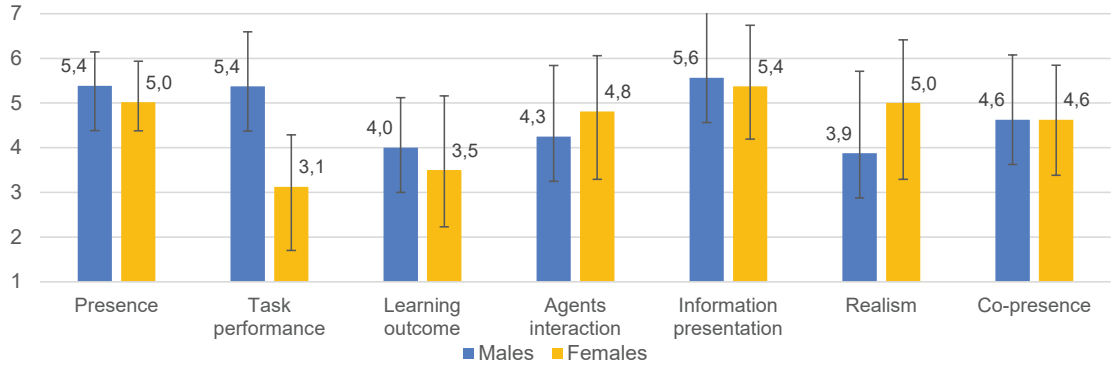

Figure 4: *The results of our questionnaire metrics in relation to participants' gender. The error bars indicate standard deviations.*

Table 2: *Significance of differences in our metrics between genders using Mann-Whitney U test. We used α value of .05.*

Metric:	Presence	Task performance	Learning outcome	Agents interaction	Information presentation	Realism	Co-presence	Duration
U value	55.5	13.5	53.5	47.5	54.5	42.5	60.5	27
p-value	.62	**.001**	.52	.32	.58	.19	.084	**.02**

Table 3: *Results of SSQ scores before and after the VR exposure. Z and p values indicate the results of the Wilcoxon signed-rank test.*

	Before VR	After VR	Z	*p*
Nausea	8.35	9.54	-.68	.5
Oculomotor disturbance	12.32	11.05	-.05	.96
Disorientation	14.5	19.72	-1.35	.18
Total simulator sickness	9.51	10.75	-1.29	.2

Table 4: *The result of open coding of participants' answers to open questions about experienced VR training, agents, and possible improvements. The left column indicates whether the codes belong to positive or negative statements expressed by participants.*

Effect	Code	Description
Positive	Realism	The environment and behaviour of agents were seen as realistic
	Speech	Conversation with an agent was rated positively and it was seen as providing useful information
	VR experience	Positive effect of VR and our scenario on training and user experience
	Gamification	Participants saw the gamification aspect of training as an advantage
	Tools	The used tools to solve incidents in the VR environment were seen as positive
	Helpfulness	Agents were seen as helpful for providing information about the environment and solving incidents in VR training
	Animation	Participants rated the animations of agents and interaction with them positively
Negative	Low realism	Some participants perceived the realism of agents and realism of conversation as insufficient
	Repeated answers	Agents responded to the same question multiple times with the same answer which seemed to be unnatural in conversation
	Insufficient emotions	Agents did not exhibit enough emotions in their behaviour and tone of voice as the participants would expect from the first response situation
	Locomotion	Some participants considered teleportation as unnatural and causing loss of orientation
	Awkward	Some participants considered agents as unresponsive, awkward, and not helpful
	One tool limit	Participants saw the limit of carrying only one tool at a time as not preferable
	Static agents	Participants would prefer to experience more active agents with higher autonomy

can see, some codes are very similar for both positive and negative effects. These ambiguities reveal different expectations of various participants with respect to individual aspects of VR first responder training and ECAs. They also highlight the importance of given codes for VR training and ECAs while the positive comments underline their benefit, the negative comments emphasize the need for improvement of these aspects in future VR training with ECAs.

5. Discussion

The main hypotheses in our study were that the Conversation condition achieves higher scores than the No-Conversation condition across the used metrics (H1-H8). Only hypothesis H7 was supported by the results of our study because only Co-presence factor exhibited statistically significant difference between conditions (Table 1). The other hypotheses were not supported. We can also see trends toward supporting some hypotheses, particularly about Agents interaction (H4) and Realism (H6). In contrast to our hypotheses, some metrics achieved lower scores in the Conversation condition than in the No-Conversation condition. These include General presence, Subjective task performance, Learning outcome, and Information presentation. Also, the duration of VR training was longer for the Conversation condition than for the No-Conversation condition. A reasonable explanation for this trend is that participants spend more time talking with the agents in the Conversation condition (because they have to ask multiple questions to discover the requested information) than just listening to their explanation in the No-Conversation condition. Therefore, the longer duration of the experiment for the Conversation condition does not necessarily mean that users are slower in performing rescue tasks. Interesting ambiguity can be observed between the trends of General Presence and Co-presence because these two metrics did not correlate in our experiment.

We also investigated gender differences in our analysis. Significant differences between genders were found in subjectively rated task performance and task duration (Table 2). In this case, males subjectively rated their task performance significantly higher than females. This finding complements previous studies about gender influence in VR on various aspects including the sense of presence [FKB*12], task performance [NG22], and others. The VR training duration was significantly shorter for males than for females in our study.

We assessed simulator sickness in our VR training by using pre-experiment and post-experiment SSQ questionnaires [KLBL93]. The highest increase in simulator sickness was 5.22 which can be categorized as minimal [BWK20]. Statistical analysis did not elicit significant differences in SSQ scores from pre-experiment and post-experiment answers.

5.1. Guidelines about ECAs in VR Training

Based on our qualitative analysis of open questions, we derive the following guidelines for future research and development of VR training applications including ECAs:

- **Realism is an important factor:** Our participants rated the realism of our VR training positively, while they also highlighted the need for improvement in this aspect. This need does not necessarily concern only visual realism but also the behavioural realism of agents with respect to a given situation including animations, actions, conversations, and emotions. Participants indicated that realistic simulation of a given training situation is important for training. E.g. the agents, involved in a car accident, do not stay in place and speak calmly but they may exhibit strong emotions and various types of active behaviour.
- **Richness of conversation is required:** While participants judged the conversation abilities of our agents as helpful, they were very sensitive in noticing conversational mistakes and unexpected conversational behaviours. Participants disliked when an agent repeated the same answer given a similar question. Additionally, participants disliked the answer of the agent to every user's statement. Therefore, the richness of the conversational model as well as smart decisions when to speak (and when to stop speaking) are desired properties of agents.
- **Situation awareness is beneficial:** The agents were seen as helpful due to providing actual information about the environment and situation-related information needed to solve the tasks.
- **Autonomy of agents is expected:** The users in our experiment expected autonomous, active, and believable behaviour of agents including not only conversational aspects but also locomotion, animations, and decisions about the next actions depending on the current situation.
- **Natural user locomotion and interaction with objects is desired:** We observed in our qualitative analysis that participants considered our teleportation locomotion metaphor as not ideal for the training scenario and they would prefer real walking locomotion. This imposes the space requirement on future experiments and application scenarios but as indicated by participants, real walking is desired to avoid disorientation caused by teleportation. Additionally, the interaction with the tools and other objects in the environment (by controllers or bare hands) is desired to be realised in a natural way (e.g. grabbing of objects, their use in solving incidents, and appropriate animations and actions of an environment with respect to the object interaction). Finally, participants requested the capability of carrying multiple items at once to increase the naturalism of the simulation.
- **Gamification is advantageous:** Participants considered the gamification aspect as advantageous for VR training. Therefore, future serious games for training purposes in VR may positively impact various training aspects including motivation and learning outcome. The positive impact of gamification on VR training was also shown in previous research [PLPK19, UBC*22].

5.2. Limitations and Future Work

While conversational agents in our VR training were considered helpful by some of our participants, there is room for improvement in future work. One of the criticised aspects of our agents was that they had a limited set of predefined answers with injected entities. Therefore, we see a high potential in using large language models for the generation of answers to improve the richness of conversation with future conversational agents.

Our agents exhibited mostly static spatial behaviour enhanced by on-place body animations. However, a realistic simulation of the first response scenario would require moving agents with advanced animations, intelligent decision-making, and emotions. Therefore, future research in agents' autonomy, agents' locomotion, and realistic emotions (both in speech and animations) is required.

As the main goal of our study was to explore the benefits of conversational capabilities of agents in a first response training, we did not evaluate learning outcome on defined pedagogic objectives and our user group was general public. This limitation might have biased the results of our study. Therefore, future studies are needed with a well-defined pedagogic scenario and expert user group from the first response domain.

6. Conclusion

In this paper, we have presented a methodology for enabling embodied agents with conversational capabilities in VR training. Additionally, the novel aspect of our agents is situation awareness. We conducted a user study to investigate the impact of the conversational capabilities of embodied agents on various aspects of VR training including presence, subjective task performance, learning outcome, interaction quality, quality of information presentation, perceived realism, co-presence, and training task duration. Our results suggest that users experience significantly higher co-presence with conversational agents than with monologue-only agents. We also discovered a significant difference in subjectively reported task performance and training duration between genders. Finally, based on our qualitative analysis, we provide guidelines for future research and development of training applications with ECAs.

Acknowledgements

We thank our participants for taking part in our user study. This research was funded by grant F77 of the Austrian Science Fund FWF (SFB "Advanced Computational Design", SP5) and by the EDIDP-SVTE-2020-047-VERTIgO project.

References

[BEGGN98] BERSOT O., EL GUEDJ P.-O., GODÉREAUX C., NUGUES P.: A conversational agent to help navigation and collaboration in virtual worlds. *Virtual Reality 3*, 1 (Mar 1998), 71–82. doi:10.1007/BF01409799. 2

[BWK20] BIMBERG P., WEISSKER T., KULIK A.: On the usage of the simulator sickness questionnaire for virtual reality research. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2020), pp. 464–467. doi:10.1109/VRW50115.2020.00098. 8

[CW19] CHETTY G., WHITE M.: Embodied Conversational Agents and Interactive Virtual Humans for Training Simulators. In *Proc. The 15th International Conference on Auditory-Visual Speech Processing* (2019), pp. 73–77. doi:10.21437/AVSP.2019-15. 1, 2

[DWW*22] DOROUDIAN S., WU Z., WANG W., GALATI A., LU A.: A study of real-time information on user behaviors during search and rescue (sar) training of firefighters. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2022), pp. 387–394. doi:10.1109/VRW55335.2022.00085. 2

[FKB*12] FELNHOFER A., KOTHGASSNER O., BEUTL L., HLAVACS H., KRYSPIN-EXNER I.: Is virtual reality made for men only? exploring gender differences in the sense of presence. In *International Society for Presence Research Annual Conference, ISPR 2012* (Philadelphia, Pennsylvania, USA, October 2012). URL: http://eprints.cs.univie.ac.at/3557/. 8

[GFOP*20] GONZALEZ-FRANCO M., OFEK E., PAN Y., ANTLEY A., STEED A., SPANLANG B., MASELLI A., BANAKOU D., PELECHANO N., ORTS-ESCOLANO S., ORVALHO V., TRUTOIU L., WOJCIK M., SANCHEZ-VIVES M. V., BAILENSON J., SLATER M., LANIER J.: The rocketbox library and the utility of freely available rigged avatars. *Frontiers in Virtual Reality 1* (2020). URL: https://www.frontiersin.org/articles/10.3389/frvir.2020.561558, doi:10.3389/frvir.2020.561558. 3

[GSMC19] GRIOL D., SANCHIS A., MOLINA J. M., CALLEJAS Z.: Developing enhanced conversational agents for social virtual worlds. *Neurocomputing 354* (2019), 27–40. Recent Advancements in Hybrid Artificial Intelligence Systems. doi:https://doi.org/10.1016/j.neucom.2018.09.099. 3

[HFR*19] HARTHOLT A., FAST E., REILLY A., WHITCUP W., LIEWER M., MOZGAI S.: Ubiquitous virtual humans: A multi-platform framework for embodied ai agents in xr. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2019), pp. 308–3084. doi:10.1109/AIVR46125.2019.00072. 3

[HZG*20] HASKINS J., ZHU B., GAINER S., HUSE W., EADARA S., BOYD B., LAIRD C., FARANTATOS J., JERALD J.: Exploring vr training for first responders. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2020), pp. 57–62. doi:10.1109/VRW50115.2020.00018. 2

[IBS11] IJAZ K., BOGDANOVYCH A., SIMOFF S.: Enhancing the believability of embodied conversational agents through environment-, self- and interaction-awareness. In *Proceedings of the Thirty-Fourth Australasian Computer Science Conference - Volume 113* (AUS, 2011), ACSC '11, Australian Computer Society, Inc., p. 107–116. 2

[JBG*19] JIN X., BIAN Y., GENG W., CHEN Y., CHU K., HU H., LIU J., SHI Y., YANG C.: Developing an agent-based virtual interview training system for college students with high shyness level. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), pp. 998–999. doi:10.1109/VR.2019.8797764. 2

[KBH*18] KIM K., BOELLING L., HAESLER S., BAILENSON J., BRUDER G., WELCH G. F.: Does a digital assistant need a body? the influence of visual embodiment and social behavior on the perception of intelligent virtual agents in ar. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (2018), pp. 105–114. doi:10.1109/ISMAR.2018.00039. 2, 3

[KdMN*20] KIM K., DE MELO C. M., NOROUZI N., BRUDER G., WELCH G. F.: Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2020), pp. 529–538. doi:10.1109/VR46266.2020.00074. 2, 3

[KLBL93] KENNEDY R. S., LANE N. E., BERBAUM K. S., LILIENTHAL M. G.: Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology 3*, 3 (1993), 203–220. doi:10.1207/s15327108ijap0303_3. 5, 6, 8

[LLGPM21] LAMBERTI F., LORENZIS F. D., GABRIELE PRATTICÒ F., MIGLIORINI M.: An immersive virtual reality platform for training cbrn operators. In *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)* (2021), pp. 133–137. doi:10.1109/COMPSAC51774.2021.00030. 2

[LPL22] LORENZIS F. D., PRATTICÒ F. G., LAMBERTI F.: Work-in-progress—blower vr: A virtual reality experience to support the training of forest firefighters. In *2022 8th International Conference of the Immersive Learning Research Network (iLRN)* (2022), pp. 1–3. doi:10.23919/iLRN55037.2022.9815975. 2

[MAB*22] MOORE N., AHMADPOUR N., BROWN M., PORONNIK P., DAVIDS J.: Designing virtual reality–based conversational agents to train clinicians in verbal de-escalation skills: Exploratory usability study. *JMIR Serious Games 10*, 3 (Jul 2022), e38669. doi:10.2196/38669. 2

[MFS*17] MOSSEL A., FROESCHL M., SCHOENAUER C., PEER A., GOELLNER J., KAUFMANN H.: Vronsite: Towards immersive training of first responder squad leaders in untethered virtual reality. In *2017 IEEE Virtual Reality (VR)* (2017), pp. 357–358. doi:10.1109/VR.2017.7892324. 2

[MMK12] MORI M., MACDORMAN K. F., KAGEKI N.: The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine 19*, 2 (2012), 98–100. doi:10.1109/MRA.2012.2192811. 3

[NG22] NENNA F., GAMBERINI L.: The influence of gaming experience, gender and other individual factors on robot teleoperations in vr. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2022), pp. 945–949. doi:10.1109/HRI53351.2022.9889669. 8

[NJD19] NGUYEN V. T., JUNG K., DANG T.: Vrescuer: A virtual reality application for disaster response training. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (2019), pp. 199–1993. doi:10.1109/AIVR46125.2019.00042. 2

[PLPK19] PALMAS F., LABODE D., PLECHER D. A., KLINKER G.: Comparison of a gamified and non-gamified virtual reality training assembly task. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)* (2019), pp. 1–8. doi:10.1109/VS-Games.2019.8864583. 8

[PSR*22] PALETTA L., SCHNEEBERGER M., REIM L., KALLUS W., PEER A., SCHÖNAUER C., PSZEIDA M., DINI A., LADSTÄTTER S., WEBER A., FEISCHL R., AUMAYR G.: Work-in-progress—digital human factors measurements in first responder virtual reality-based skill training. In *2022 8th International Conference of the Immersive Learning Research Network (iLRN)* (2022), pp. 1–3. doi:10.23919/iLRN55037.2022.9815976. 2

[PSSE21] PERETTI O., SPYRIDIS Y., SESIS A., EFSTATHOPOULOS G.: Gamified first responder training solution in virtual reality. In *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)* (2021), pp. 295–301. doi:10.1109/DCOSS52077.2021.00055. 2

[RHW20] REINHARDT J., HILLEN L., WOLF K.: Embedding conversational agents into ar: Invisible or with a realistic human body? In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction* (New York, NY, USA, 2020), TEI '20, Association for Computing Machinery, p. 299–310. doi:10.1145/3374920.3374956. 1, 3

[SBS19] SCHMIDT S., BRUDER G., STEINICKE F.: Effects of virtual agent and object representation on experiencing exhibited artifacts. *Comput. Graph. 83*, C (oct 2019), 1–10. doi:10.1016/j.cag.2019.06.002. 1, 2

[SLB*23] SCHLESENER E. A., LANCASTER C. M., BARWULOR C., MURMU C., SCHULENBERG K.: Titleix: Step up & step in! a mobile augmented reality game featuring interactive embodied conversational agents for sexual assault bystander intervention training on us college campuses. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2023), CHI EA '23, Association for Computing Machinery. doi:10.1145/3544549.3583832. 3

[SNS19] SCHMIDT S., NUNEZ O. J. A., STEINICKE F.: Blended agents: Manipulation of physical objects within mixed reality environments and beyond. In *Symposium on Spatial User Interaction* (New York, NY, USA, 2019), SUI '19, Association for Computing Machinery. doi:10.1145/3357251.3357591. 3

[SRD15] SHARMA S., RAJEEV S. P., DEVEARUX P.: An immersive collaborative virtual environment of a university campus for performing virtual campus evacuation drills and tours for campus safety. In *2015 International Conference on Collaboration Technologies and Systems (CTS)* (2015), pp. 84–89. doi:10.1109/CTS.2015.7210404. 2

[SSS98] STANSFIELD S., SHAWVER D., SOBEL A.: Medisim: a prototype vr system for training medical first responders. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No.98CB36180)* (1998), pp. 198–205. doi:10.1109/VRAIS.1998.658490. 2

[TR02] TRAUM D., RICKEL J.: Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 2* (New York, NY, USA, 2002), AAMAS '02, Association for Computing Machinery, p. 766–773. doi:10.1145/544862.544922. 3

[TRO*19] TECHASARNTIKUL N., RATSAMEE P., ORLOSKY J., MASHITA T., URANISHI Y., KIYOKAWA K., TAKEMURA H.: Evaluation of Embodied Agent Positioning and Moving Interfaces for an AR Virtual Guide. In *ICAT-EGVE 2019 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments* (2019), Kakehi Y., Hiyama A., (Eds.), The Eurographics Association. doi:10.2312/egve.20191276. 3

[UBC*22] ULMER J., BRAUN S., CHENG C.-T., DOWEY S., WOLLERT J.: Gamification of virtual reality assembly training: Effects of a combined point and level system on motivation and training results. *International Journal of Human-Computer Studies 165* (2022), 102854. doi:https://doi.org/10.1016/j.ijhcs.2022.102854. 8

[VRAG*14] VÉLAZ Y., RODRÍGUEZ ARCE J., GUTIÉRREZ T., LOZANO-RODERO A., SUESCUN A.: The Influence of Interaction Technology on the Learning of Assembly Tasks Using Virtual Reality. *Journal of Computing and Information Science in Engineering 14*, 4 (10 2014). doi:10.1115/1.4028588. 2

[VWG*04] VORDERER P., WIRTH W., GOUVEIA F., BIOCCA F., SAARI T., JÄNCKE L., BÖCKING S., SCHRAMM H., GYSBERS A., HARTMANN T., KLIMMT C., LAARNI J., RAVAJA N., SACAU A., BAUMGARTNER T., JÄNCKE P.: *MEC spatial presence questionnaire (MEC-SPQ): Short documentation and instructions for application*. Tech. rep., Report to the European Community, Project Presence: MEC (IST-2001-37661), 06 2004. 5

[WS98] WITMER B. G., SINGER M. J.: Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments 7*, 3 (06 1998), 225–240. doi:10.1162/105474698565686. 5

[WSR19] WANG I., SMITH J., RUIZ J.: Exploring virtual agents for augmented reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, p. 1–12. doi:10.1145/3290605.3300511. 2, 3