Audio vs. Visual Avatars as Guides in Virtual Environments

Nico Brettschneider and Jens Herder
Hochschule Düsseldorf - University of Applied Sciences
Münsterstraße 156
40476 Düsseldorf, Germany
nico.brettschneider@gmail.com
jens.herder@hs-duesseldorf.de

Jeroen de Mooij
Weird Reality
Kaiserin-Augusta-Allee 101
10553 Berlin, Germany
Jeroen@thefirstfloor.nl

Bektur Ryskeldiev
Digital Nature Group
University of Tsukuba
Tsukuba, Ibaraki, Japan
bektour@digitalnature.slis.tsukuba.ac.jp

***Abstract*—Through constant technical progress, multi-user virtual reality is transforming towards a social activity that is no longer only used by remote users, but also in large-scale location-based experiences. We evaluate the usage of realtime-tracked avatars in co-located business-oriented applications in a "guide-user-scenario" in comparison to audio only instructions. The present study examined the effect of an avatar-guide on the user-related factors of *Spatial Presence, Social Presence, User Experience* and *Task Load* in order to propose design guidelines for co-located collaborative immersive virtual environments. Therefore, an application was developed and a user study with 40 participants was conducted in order to compare both guiding techniques of a realtime-tracked avatar guide and a non-visualised guide with otherwise constant conditions. Results reveal that the avatar-guide enhanced and stimulated communicative processes while facilitating interaction possibilities and creating a higher sense of mental immersion for users. Furthermore, the avatar-guide appeared to make the storyline more engaging and exciting while helping users adapt to the medium of virtual reality. Even though no assertion could be made concerning the *Task Load* factor, the avatar-guide achieved a higher subjective value on *User Experience*. Due to the results, avatars can be considered valuable social elements in the design of future co-located collaborative virtual environments.**

***Index Terms*—Virtual Reality, Co-located Collaborations, Networked Immersive Virtual Environments, Head-mounted Display, Avatars, Social Presence**

I. Introduction

In a globally connected society, the sense of "being there together" still has a strong impact on people's emotions. In Virtual Reality (VR), this sense may also be there, but is still limited by technical implementations. Since VR has been growing over the past years, the idea of having multiple people in the same VR experience is not new but is constantly getting refined and adapted to the latest technology. While networked solutions are already available for consumers in form of *social virtual reality*, large-scale location-based multi-user systems are also on the rise to a mass-consumer market [4]. The usage of avatars in social VR appears to be obvious, as remote users need to be visually represented. Furthermore, the "social significance" of avatars [21] is being discussed in related literature for almost two decades already. But how about co-located multi-user VR? Rather related to the business- than to the consumer-sector, location-based collaborative VR provides a variety of possibilities regarding i.e. virtual product presentation or industrial trainings with a local group of users. Such "guide- user-scenario" use cases, where an expert, trainer or salesman is together with users, raise the question of whether a realtime-tracked visualised guide is beneficial for the highly success-related user's personal experience. Social and spatial influences on users of a realtime-tracked avatar-guide were not yet scientifically evaluated in this specific scenario. Therefore, this article addresses commonly used user-oriented factors in this field of research, aiming to propose design guidelines for co-located collaborative immersive virtual environments (IVEs), by answering the following question:

> Does a realtime-tracked avatar-guide in a co-located collaborative IVE enhance Spatial Presence, Social Presence, User Experience and Task Load for users?

In order to be able to answer that question, a comparison has to be made between the realtime-tracked avatar-guide and a non-visualised-guide. This is done by developing a co-located collaborative IVE application based on related research and performing a user study which utilizes commonly-used questionnaires to obtain subjective data. Besides that, objective data is recorded on task performances. For the most reliable outcome the same person performs the role as guide in both guiding techniques.

II. Related Work

In this paper we expand on experiment setup and methods briefly introduced in [9].

A. Avatars in Immersive Virtual Environments

As long ago as 2000 Slater et al. [21] investigated on small group behaviour and performing a task in a shared virtual environment rather from a technical, than from a social point of view. The main results of their experimental study were that "personal responses to social situations [...]" are also reclaimable in a shared VE, even though the visual

Paper presented at HC-2018, March 27–28, 2019, Shizuoka University, Hamamatsu, Japan.

representation of each other is limited. The limited avatars appeared to have a "social significance", as people tended to respect each other's avatars. Besides they noticed a positive association between presence and co-presence, which later got apprehended by Casanueva and Blake [18], who though did not find a positive correlation between the two factors. Nevertheless they ascertained, that " [...] avatars having gestures and facial expressions produced a significantly higher level of co-presence when compared to static avatars.", but also, that human-like avatars with a higher state of realism but regardless to gestures and facial expressions, created a higher sense of co-presence compared to unrealistic avatars without body movement.

A recent study about "Avatar Realism and Social Interaction Quality in Virtual Reality" by Roth et al [16] confirms the impeding tendency of non-realistic avatars on social interaction, but points out, that the lack behavioural cues like gaze and facial expression can partly be compensated. Additionally the effectiveness in a communicative role-play which compared a verbal task in VR to the same in the real world, did not differ. Lugrin et al. [13] also found in a comparing experiment, that user performance and user experience is to a certain degree "not degraded by abstract or iconic visual representations", while self-reported presence is always higher with avatar embodiment independent from its graphic fidelity. In the study of Heidicker et al. [8] a fully mapped complete avatar body caused the highest level of co-presence, but an avatar only consisting of head and hands was not significantly worse.

In a comparing study by Costigan et al. [2], a remote instructor in an immersive collaborative virtual learning environment was visualized by a video window, as well as by a computer generated avatar. Comparing the two methods with an "in-person" condition which implies the instructor sharing the same physical space, significant similarities between avatar- and in-person condition were found, while the video condition lacked behind the two. Furthermore it showed that audio in combination with "physical movement" were rather important for representing a persona than a "visually correct image transmitted via video". The real-time audio indicated itself as a "primary need for creation and maintenance of the real-time feel of the environment" and it was "creating the telepresence", even when visual errors such as crashes or delays occurred. By investigating in the impact of avatar eye gaze, Steptoe et al. [22] compared tracked eye gaze to static eye gaze and model-based eye gaze which took objects and other avatars in a user's field of view randomly into account. Findings showed, that the model-based eye gaze significantly lacked behind the two others in quality of communication while its incorrectness showed up as a "hindrance during object-focused interaction in ICVEs". In an experiment with photorealistic avatars by Latoschik et al. [11], the *Illusion of Virtual Body Ownership (IVBO)* was measured higher with realistic avatars compared to wooden mannequin avatars. Furthermore the appearance of another avatar has an impact on the users "self-perception towards their own body". Investigating further in the impact of a "self-avatar" in a SVE and using *HTC Vive* consumer VR devices, Pan and Steed [14] measured quicker task completion times in a collaborative task with two users having a self-avatar compared to two users just having visual representations of controllers. Moreover they analysed a higher value of self-reported "trust" between the participant couples having a self-avatar. In a medical experiment of Russo et al. [17] the impact of having a shadow in VEs is examined in a pilot study with post stroke inpatients. It resulted an improvement of "sustained attention" due to having a body shadow, which apparently puts attention also to the body itself.

B. *Collaborative Virtual Environments*

Real time co-located interaction in VR is, according to the findings of Greenwald et al. [5] a "practical medium for communication and collaboration" whilst carrying with "a sense of social presence". In a very current study by Wienrich et al. [24], "social interdependence" in a room-scale location-based VR experience had a beneficial effect on social presence, cooperation and affective evaluation of the users. Also using a room-scale collaborative setup Lacoche et al. [10] pointed out the "safety of users" aspect and designed different user representation types for warning purposes in shared physical spaces. Proposing three different visualization methods, two stood out after the study concerning "safety of users" and "global satisfaction". The "ghost avatar", a visualisation of HMD model and controllers, and the "extended grid", a cylinder-shaped grid around the user, appeared both, when another user got physically close to a user. Due to the positive results, both could be well working alternatives instead of separating the tracked space. To figure out the suitability of HMDs for collaborative visualization, Cordeil et al. [1] compared a collaborative CAVE-style environment to an HMD-based one using *Oculus Rift* HMDs. Comparing the differences between the two VR platforms, the subjective and objective measures for the two 3D visualized network tasks concerned "accuracy", "times", "movements", "oral communication", "task allocation" and "strategies". As a key finding they figured out, in both environments participants where highly accurate, but also substantially faster with the HMD. Furthermore they found no major differences in terms of oral communication. By investigating the utility of co-located collaborative interaction techniques in a projection- and HMD-based-setup Salzmann et al. [19] confirmed, that individual perspective correctness of the virtual world is essentially required for collaborative work.

Summing up the analysis of related work concerning avatar usage in IVEs indicate a positive effect on users in terms of presence, co-presence and user experience. However avatar realism seems to be limitable to a certain extend whilst not significantly affecting this effect. Furthermore the topic of CVEs is very current, even the technical implementation tends to change into HMD-based setups. For co-located setups, spatial problems have been analysed and influences of visual representation of others has been found in different experiments. The latter is, in combination with audio, highly necessary for communicating and also guiding processes.

III. METHOD

A. Participants

Forty volunteers (n = 19 female, M = 29.43 years; SD = 8.13 years) with normal (n = 24) or corrected vision participated in the study. In a pre-questionnaire participants were asked besides demographic data about their VR and 3D expertise with several questions (5 point Likert-scale), while "0" would be "none" and "4" would be "very high" knowledge or expertise. Answers to specific VR-usage-frequency questions showed that more than half of the participants (n=24) have used stationary VR devices and more than a third (n=14) have used mobile VR devices once of more times before. Additionally all answer points were summed up to estimate an "3D and/or VR Expertise" score. When a participant reached the hypothetical score of "8/32", a score he would reach by checking every question with one point ("little"), he is considered "Experienced", similar to the approach of Tcha Tokey et al. [23]. 17 participants were considered "Experienced" in the experiment. No compensation has been paid for participation in the study.

B. Materials

1) Hardware and Software Installation: Due to a remarkable amount of usage in related literature, but also focussing on the fact of a five square meter room-scale tracking capability, the decision has been made to use *HTC VIVE* (HTC, Corp.) HMDs and Controllers for the VR application. The measured play area thereby had a space of 3.4 by 3.4 meters. The server space was separated from the participant area space with soundproof glass. Three Microsoft Windows-based computers (2 Clients, 1 Server) were used to run the software build. The room setup has been done on one computer and then got shared to the other to match the boundaries and room settings, while the server computer had no VR devices attached. The virtual environment has been created in Unity3D (Unity Technologies), furthermore models were created in Cinema4D (MAXON, GmbH). The audio setup has been realized with a custom 4.1 speaker setup powered by a *RME Fireface UC Audio Interface* (Audio, AG) including a custom channel routing. The experimenter in the operator/server space was able to communicate with the participant in the play area space via a talk back microphone at any time and vice versa. Software-wise the whole VR experience is split up in "Stages". A "Stage" is a custom term declaring a certain phase or chapter in the experience, whereby a "Stage" is always connected with a Unity scene. Within the experience, a guide is able to change the current "scene" at any time for all users. Furthermore he is able to show "Safe Spots" on the ground which mark the initial position in the room for each respective user. Due to large scale stages, group teleportation was also implemented. Furthermore, critical story elements and interactable parts could be highlighted from the guide. A necessary part for comparing the two guiding techniques of a realtime-tracked avatar-guide (further referred to as “avatar-guide”) and a non-visualised-guide (further referred to as “audio-guide”) is keeping the interaction possibilities of the guide with the user as equal as possible. Therefore all guide functionalities were also implemented in a server canvas. As the experience depends on solving tasks, a "Guide Assistant" system with a "Timer" and countdown was implemented and displayed to the Guide’s HUD and server canvas, keeping the timing of giving task-related hints to the user as constant as possible.

2) Questionnaires and Measurements: All Questionnaires have been assessed in digital form via *Google Forms* (Google, LLC) to ensure a quick assessment before and after each study run. Before the participants were able to take part in the study, everyone was asked to fill in a “pre-questionnaire” collecting the “experience level” of each participant. After the practical part of the study was over, three subjective key measurements were be performed in the following order with a "post-questionnaire" consisting of commonly used questionnaires:

- Sense of spatial- and social presence
- Task load
- User Experience

The first measurement *Sense of spatial- and social presence* has been performed based on the "Temple Presence Inventory (TPI)" [12], which especially takes a "social component" into account. The subscales of the original 42 seven-point Likert-scale item questionnaire used were "Spatial Presence (N=7)", "Social Actor w/i Medium (N=6)", "Social Passive Interpersonal (N=4)", "Social Active Interpersonal (N=3)" and "Engagement (N=6)", while the other subscales were dropped due to a lack of relevance or doubling with the User Experience measurement, which was performed with a separate questionnaire.

To measure *Task Load*, the "NASA TLX" [7] was used due to its establishment in related literature and high comparability. As Gross [6] summed up the negative aspects of a preliminary weighting process in the original version of the TLX, the questionnaire can be evaluated without weighting. The values of the six subscales have therefore been added and divided by the count of the subscales to ensure comparability.

User Experience was measured with the "User Experience Questionnaire (UEQ)" [20] [1] including the subscales "Attractiveness", "Perspicuity", "Efficiency", "Dependability", "Simulation", "Novelty". Compared to other commonly used UX questionnaires, it advantages also with measuring the complete UX with not just technical but also visual aspects and a personal opinion.

In addition to the subjective measurements, times and errors have been logged within the system for each participant and task. Therefore, the "Start Time" on enabling interactable elements for the specific task, as well as the "End Time" after completing the last part of an interactive exercise was saved. Individual "Task Completion Times" were calculated afterwards with the logged data. For the errors, an "Error" score was calculated by summing up every mistake or nonconformity with the Guide’s instructions that has been

[1] https://www.ueq-online.org/

done in an individual exercise, i.e. pushing a wrong button. Additional behavioural observations were performed by one of the experimenters during each experimental run.

C. Procedure

1) Experimental Session Structure: The experimental sessions were split up into *pre-study phase*, *during-study phase* and *post-study phase*. For every participant, an overall session time of 30 minutes was estimated, including all phases. Initially in the *pre-study phase* every participant was asked to arrive 5 minutes earlier than his session time, to get a short introduction and safety advises. Furthermore the participants' informed consent was retrieved and the pre-questionnaire was filled in if not happened already, as it was sent with the invitation email before. In order to minimize external influences, participants were asked to stay outside the testing room until immediately before their experimental run. As the experiment was set up as a "Between Subject Design", Participants got no information about their experimental group or the purpose of the experiment to prevent affected results. Subsequently participants were called independently to go inside to get an introduction in the VR equipment by one of the experimenters. The experience (*during-study phase*) started with an orientational "Personal Space". If the participant was in the "AVATAR" study group, the guide person, one of the experimenters, entered the room not visible for the participant and took his position. After the participant pressed a start button, the space changed to a "Lobby", where the guide was introduced.

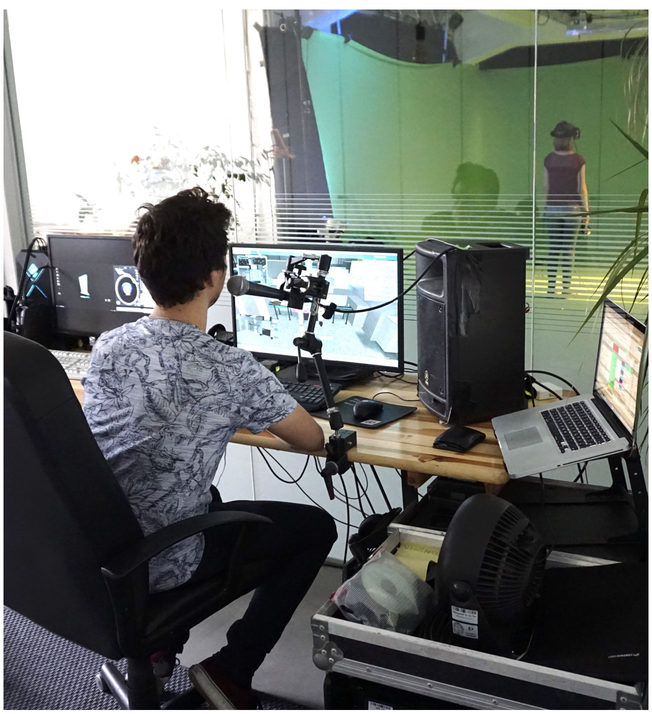

Fig. 1. Audio-guide with one participant (view from operator/server space)

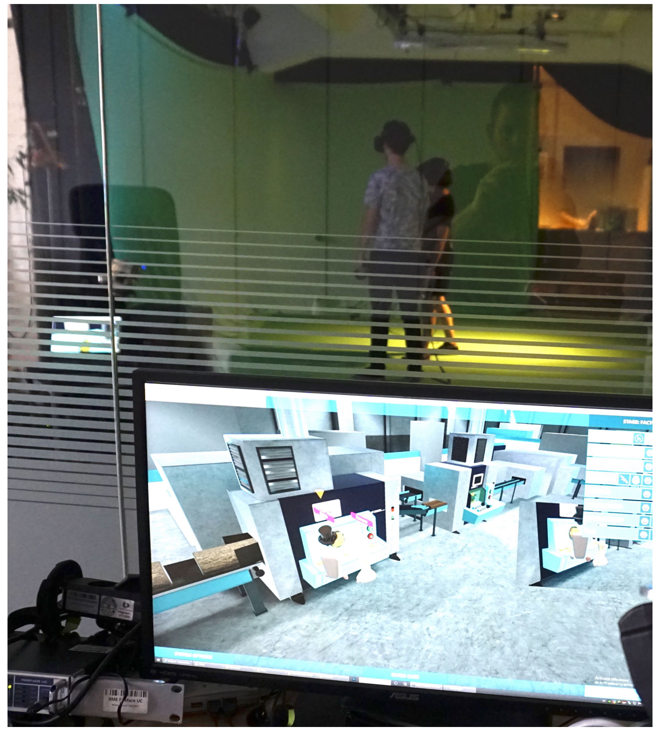

Fig. 2. Avatar-guide with two people co-located (view from operator/server space)

Depending on the study group, the guide introduced himself verbally via the speakers in the room (see Figure 1) or verbally and visualised by an avatar (see Figure 2). Moreover, the audio-guide communicated over the room speakers, as the spatial audio information of his position in the same room might have a confusing effect on the participants. The guide introduced the teleporting system and changed the “Space” to the “Factory”, where the actual tasks were performed. After finishing the "Factory", the scene changed to an "End Space", where the participant was thanked, instructed to take off the HMD and give it to one of the experimenters. The *post-study phase* phase began and the participant was asked to fill in the post-questionnaire immediately.

2) Storyline and Tasks: The overall story of the experience is divided into a "framework part" including the whole experience with different "Stages" and a "actual-story part" which only includes a so called "Factory Stage". The "framework part" starts in the "Personal Space", where the user gets introduced into the virtual world and sees himself as an avatar in a virtual mirror. After getting in contact with the guide in the "Lobby Stage", where the user also got introduced in the necessary interactions with the VIVE controllers, the "Factory Stage" as "actual-story part" of the storyline begins. The "Factory" is a large modern factory hall with futuristic-seeming machines. Tree trunks on a conveyor belt and pallets of identical plain wooden chairs indicate the wood processing function of the assembly line. The guide explains the factory and the machines. He explicitly points at or highlights the result of the machines in form of the pallets of chairs and a big "prototype chair", as participants are the new factory workers and need to know what is produced. The guide explains, that eventually an electric malfunction has occurred and teleports with the participant to the "Power Machine", where the participant needs to restore the power supply by switching on five fuses. After restoring the power, the machines indicate the restored power, but still have a malfunction as they "deleted" their programs and need to be overwritten manually. The guide again mentions to look at the prototype chair, as the final product after overriding the machines manually should look exactly like the ones on the pallets or the prototype representation. Proceeding the guide is teleporting with the participant to each machine in the assembly line producing a respective part of the chair.

The tasks at each machine differ slightly from each other and consist of pushing buttons and sliders (see Figure 3). Every task requires logical thinking, while the guide is only giving technical hints to solve the task and a monitor attached to the respective machine gives a hint to the outcome of the machine. After finishing the last task, a chair comes out of the machine and the guide asks the participant to pick it up and compare it to the ones on the pallets. Moreover, the participant gets confronted with the mistakes he/she made during the production process. After a positive comment of the guide on the resulting chair, no matter how the actual result looks like, the guide changes the Stage to the "Ending Stage". A few seconds later the chairs fall down like stars and a message appears to take off the HMD, as the experience ends.

IV. Results

A. Filtering

Due to technical issues with the logging system, the sample size for the log-file-analysis was reduced to N=36 (AVATAR-Group: N=16, AUDIO-Group: N=20). The UEQ was analysed with the integrated analysis tool comparative with both experimental group as well as individually for each group. With the heuristic test detecting suspicious data in terms of randomness or lacks of seriousness by checking the difference of maxima and minima per subscale with a limit value ($Diff(Max, Min) > 3$), the sample size was also reduced to N=33 (AVATAR: N=17, AUDIO: N=16). The Test was also applied to the TPI questionnaire with $Diff(Max, Min) > 4$ not considering the "eyecont" question as it was visual specific and therefore obviously biassing the test. Hence the sample size again was reduced to N=36 (AVATAR: N=18, AUDIO: N=18). For the NASA TLX, no participant was excluded.

B. Questionnaire Results

At first, reliability was tested with Cronbach's Alpha [3] for the TPI and UEQ questionnaires (see Table 1). For the two independent samples, significances were found with the non-parametric “Mann-Whitney-U” rank sum test [15]. Assuming greater means with the AVATAR group, a one-tailed test provided significant results for the TPU items in the subscales listed in Table 2. No further significant difference between the two groups were found for the UEQ and NASA TLX questionnaires.

Subscale	Cronbach's Alpha for AVATAR-Group	Cronbach's Alpha for AUDIO-Group
Spatial Presence (TPI)	0.55	0.58
Social Actor w/i Medium (TPI)	**0.77**	**0.87**
Social Passive Interpersonal (TPI)	**0.87**	**0.84**
Social Active Interpersonal (TPI)	**0.84**	**0.89**
Engagement (TPI)	**0.81**	**0.79**
Attractiveness (UEQ)	**0.84**	**0.83**
Perspicuity (UEQ)	**0.81**	**0.89**
Efficiency (UEQ)	0.37	**0.70**
Dependability (UEQ)	**0.72**	0.61
Stimulation (UEQ)	**0.73**	0.55
Novelty (UEQ)	0.45	0.59

TABLE I
CRONBACH'S ALPHAS [3] FOR TPI AND UEQ SUBSCALES PER GROUP; **BOLD** TYPE CONSIDERED HIGHER THAN A COMMONLY ADEQUATE LEVEL OF 0.7

Item (TPI Subscale)	U-value	p-value
mksound (Social Active)	98	$< .05$
smile (Social Active)	66	$< .01$
speak (Social Active)	85.5	$< .01$
contrint (Social Actor)	102.5	$< .05$
eyecont (Social Actor)	46	$< .01$
interact (Social Actor)	101	$< .05$
leftplce (Social Actor)	96.5	$< .05$
bodylang (Social Passive)	46.5	$< .01$
faceexpr (Social Passive)	97.5	$< .05$
styledres (Social Passive)	77	$< .01$
exciting (Engagement)	105.5	$< .05$

TABLE II
MANN-WHITNEY-U SIGNIFICANT RESULTS FOR TPI

The AVATAR-group reached higher values in terms of Social Presence, while results of the TPI Presence Inventory are almost throughout positive for both groups. Especially for the subscale "Social Active Interpersonal" (see Figure 4)) the AVATAR group reached clearly higher means for the three items "How often did you make a sound out loud [...]" (mksound), "How often did you smile in response to someone [...]" (smile) and "How often did you want or did speak to a person [...]" (speak).

Furthermore, the subscale aiming at "Parasocial Interaction" named Social Actor W/I Medium also showed diverse item

Fig. 3. Exemplary machine with sliders highlighted by audio-guide (with no visual avatar)

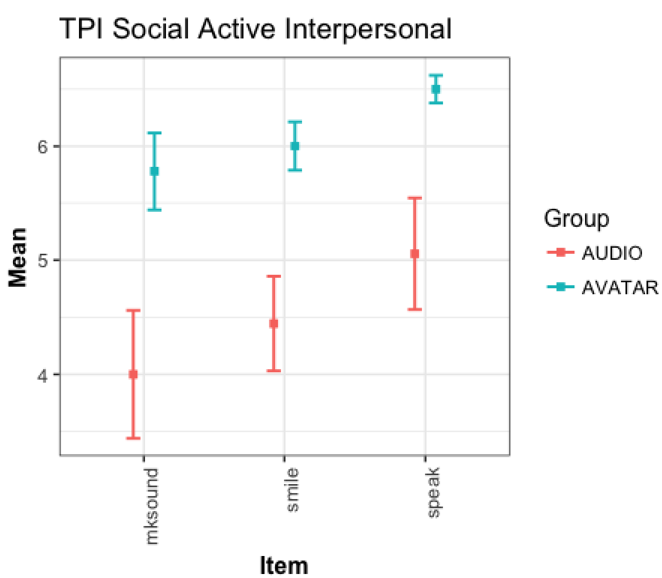

Fig. 4. Means and standard errors for Social Active Interpersonal subscale

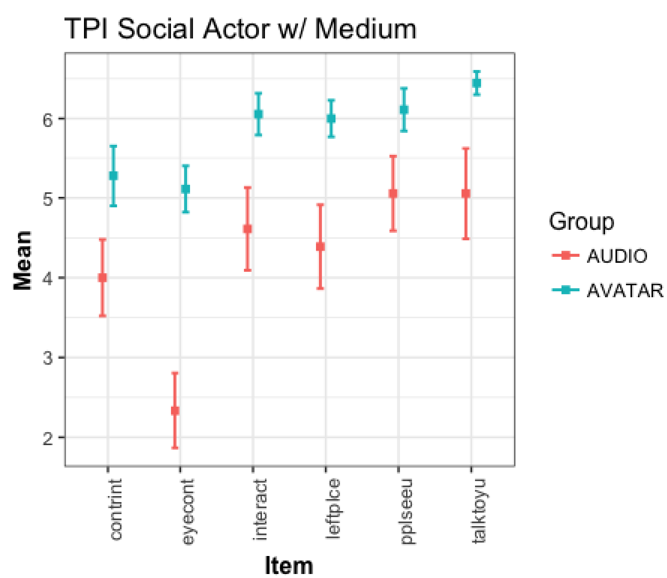

Fig. 5. Means and standard errors for Social Actor W/I Medium subscale

means between the two groups (see Figure 5). AVATAR group participants reported a higher "control over the interaction with the person they saw/heard" (contrint), a higher feeling of being able to "interact with the person they saw/heard" (interact) as well as a higher "sense of leaving the place they were" (leftplce). Moreover, the items "How often did you want or did you make eye contact" (eyecont), "How often did you have the sensation that people you saw/heard could also see/hear you" (pplseeu) and "How often did it feel as if someone you saw/heard in the environment was talking directly to you" (talktoyu) were also rated higher than in the AUDIO group.

Besides social presence, the subscale "Engagement" aiming at "mental immersion" reached higher means with the AVATAR group for almost every item (see Figure6)).

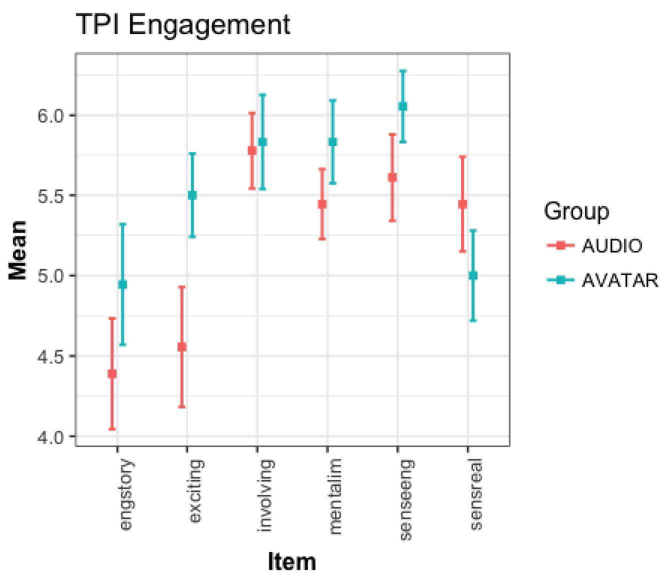

Fig. 6. Means and standard errors for Engagement subscale

While the experience in the AVATAR group tended to be more "engaging" (engstory) and "exciting" (exciting), the "involving" (involving) factor appeared to be almost equal for both test groups. The AVATAR group also reported a higher sense of "mental immersion" (mentalim) as well as a higher feeling of "engaged senses" (senseeng). Merely the AUDIO group stated a higher "sensation of reality" (sensereal) within the experience. Even though the UEQ showed no significant differences, both groups reached high values for UX in the subscale "Attractiveness" and "Perspicuity", while the AVATAR groups was rated a little better than the AUDIO group (see Figure 7). The subscales "Efficiency" and "Novelty" reached similar means, while the AUDIO group has a little advance in means for "Dependability" and "Stimulation". Concerning the Nasa TLX, the "Mean Task Load" was M = 23.88, SD = 11.16 for the AVATAR group and M = 21.79, SD = 13.12 for the AUDIO group. One-tailed Mann-Whitney-U tests showed no significant difference in means for the single items and the "Mean Task Load". Moreover, assuming no difference in means for "Experienced" and "Non-experienced" participants an in-group two-tailed Mann-Whitney- U test proved the assumption for all items and "Mean Task Load".

C. Additional Results

The whole experience took 9 minutes and 58 seconds on average (AVATAR: M = 09:46, AUDIO: M = 10:07) including the framework storyline. The actual interactive story part ("Factory Scene") took averagely 8 minutes and 13 seconds (AVATAR: M = 08:00, AUDIO: M = 08:13). Task completion times were almost equal for both groups per machine (mean difference between groups per machine: 1.5 seconds) and therefore not significantly different. Experienced participants tended to solve almost all tasks slightly quicker in both groups. By looking at the errors or failures in the tasks, the AUDIO group apparently made more mistakes (M = 2.45, SD = 1.23) than the AVATAR group (M = 1.94, SD = 0.85). An "Error" is defined as a mistake in one of the three result-affecting machines and results showed that no participant created an entirely "correct chair". In-group comparison of experienced and non-experienced participants resulted in no general assertation for the single tasks, as the difference in mean was either slightly alternating within the groups of just was barely noticeable. Regarding the total amount of errors, a higher tendency resulted for the non-experinced participants (AVATAR: M = 2, SD = 1 / AUDIO: M = 2.69, SD = 1.18) compared to the experienced participants (AVATAR: M = 1.86, SD = 0.69 / AUDIO: M = 2, SD = 1.29). Furthermore, errors were compared by participants used to wear glasses and participants without glasses. Results showed a higher error rate in total for glass-participants (M = 2.57, SD = 1.09) contrasting with no-glass-participants (M = 2, SD = 1.07) with the group allocation of AVATAR: N(glasses) = 4 and AUDIO: N(glasses) = 10.

V. Discussion

A. Main Results

Initially and in line with the expectations derived from related work, the study showed significantly higher results for the AVATAR group in terms of "Social Presence" as the TPI showed. Especially active interpersonal interactions showed up to be significantly stimulated by an avatar-represented guide in a location-based space. This positive effect on participants appeared not only in the will of reacting to the guide by making a sound or smiling as a response, but also as a trigger for self-induced verbal conversation. In this connection the avatar can be interpreted as a catalyst for active social interaction. Not only interpersonally but also through the medium of the virtual environment, the avatar-guide tends to be significantly beneficial compared to the audio-based one, as the avatar seems to have a relieving influence on the control of interaction, as well as on interactions themselves. Beyond that the feeling of "leaving the place you currently are and going to another place" was also significantly higher with an avatar-guide. Therefore, he appears as a massive intensifier for the feeling of "being somewhere new" while he provides an audio-visual communication channel as a tracked person compared to just audio communication. As the better control of interaction and interaction possibilities with the avatar could also carry a certain feeling of security for a user, this might be the reason for the significantly better active interpersonal interaction the AVATAR group showed. The question of whether people wanted to make or did make eye contact with the people they saw/heard was rated obviously better with the AVATAR group, as the other group did not have a visually represented second person in the virtual environment. As the difference is extraordinary, the AUDIO group likely understood rather as a

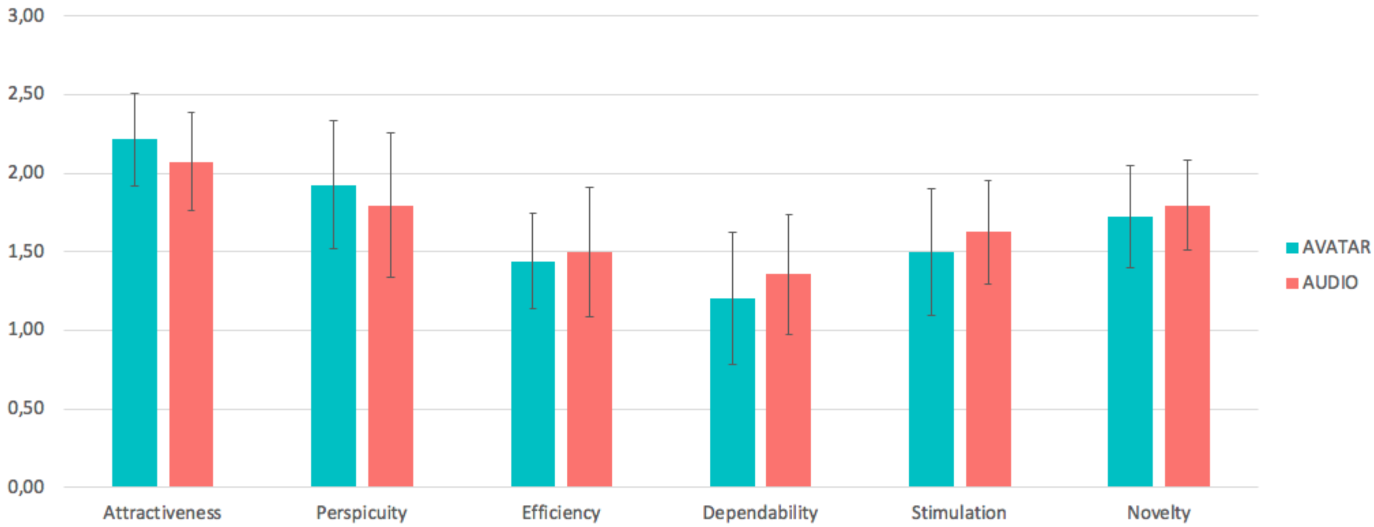

Fig. 7. User Experience Questionnaire (UEQ) scale means with error bars for the AVATAR and AUDIO group

question of the actual interaction of making eye contact than the will of making eye contact with someone. Another explanation would be that the lack of a visual representation decreases the need of making eye contact in a conversation. The AVATAR group also achieved better results in the questions asking if people participants saw/heard could see/hear/talk to them which again can be attributable to a misunderstanding of the question. The AUDIO group obviously cannot see anyone but they assumable understood the questions only in that particular way, even though the audio-guide saw them all the time through the window and was talking directly to them over the speakers. It is also possible, that the fact of only hearing someone affects the self-perception of being seen or heard by others. According to engagement, the overall story tended to be more engaging and was significantly more exciting for AVATAR group participants. As the story was entirely identical and the only difference was the absence of an avatar for the guide in the AVATAR groups, it stands to reason that the avatar not only has an influence on social interaction but also enhances the presented story, surrounding and exercises subjectively. The deliberately banal storyline gets evidently better through the simple presence of the other person being the guide in the same room as an avatar. This hypothesis gets supported by the results for the question of how "involving" the story was. Because both groups reached nearly the same means, one can assume that the participants knew and understood their role in the story for both groups equally as they reported an almost equal level of involvement. However feeling involved in a story appears to not necessarily accompany with the senses of engagement and excitement, as the group comparison showed. Results for the items concerning "mental immersion" and the "engagement of senses" also showed positive tendencies for the AVATAR group. Mental immersion could firstly be influenced by the stronger social interaction possibility, as related work showed before. Both items likely also go hand in hand with the level of engagement and excitement, whereas an exciting and engaging story could lead to stronger focus and thus a higher mental immersive state. Concerning the increasingly engaged senses, an avatar as a visual appearance in the virtual environment the AUDIO group did not have can possibly be responsible for the latter. Interestingly the AVATAR group lacked behind the other in terms of a "sense of reality". This could be a result of the very limited comic-style avatar design, which supports the impression of a "game-style" environment rather than a serious educational surrounding. By contrast the AUDIO group could have interpreted the environment as "more realistic", as a remote instructor giving instructions over speakers in the factory hall is not a completely unlikely scenario. Furthermore, the AUDIO group may had a stronger visual impression of the virtual environment while not being distracted by the avatar-guide. Although the tasks were still far away from actual realistic machine interactions, the focus on the surrounding while getting instructions from "outside" and being observed could lead to those results.

VI. Conclusion

We have shown that simple avatars controlled only by HMD and controller can be considered as valuable social elements in the design of co-located collaborative interactive virtual environments. Future developments will enhance tracking and visual possibilities, so that avatars become a common element of a VR experience.

Acknowledgments

The experiment was set up in the premises of Burmester Event- und Medientechnik GmbH, who also provided all technical equipment.

References

[1] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott, and B. H. Thomas. Immersive Collaborative Analysis of Network Connectivity: CAVE-style or Head-Mounted Display? *IEEE Transactions on Visualization and Computer Graphics*, 23(1):441–450, Jan. 2017. doi: 10.1109/TVCG.2016.2599107

[2] J. T. Costigan, A. Johnson, and S. Jones. Comparison of Remote User Representation in a Collaborative Virtual Learning Environment. *Journal of Virtual Environments*, 6(1), Jan. 2002.

[3] L. J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334, Sept. 1951. doi: 10.1007/BF02310555

[4] C. Fink. The Future Of Virtual Reality Isn't Your Living Room - It's The Mall, Mar. 2017.

[5] S. W. Greenwald, Z. Wang, M. Funk, and P. Maes. Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality. In D. Beck, C. Allison, L. Morgado, J. Pirker, F. Khosmood, J. Richter, and C. Gütl, eds., *Immersive Learning Research Network*, vol. 725, pp. 75–90. Springer International Publishing, Cham, 2017. doi: 10.1007/978-3-319-60633-0_7

[6] B. Gross. *Bestimmung von Schwierigkeitsgraden in einer zu entwickelnden Versuchsumgebung*. Diplomarbeit, Humboldt-Universität zu Berlin, Berlin, 2004.

[7] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock and N. Meshkati, eds., *Advances in Psychology*, vol. 52 of *Human Mental Workload*, pp. 139–183. North-Holland, Jan. 1988. doi: 10.1016/S0166-4115(08)62386-9

[8] P. Heidicker, E. Langbehn, and F. Steinicke. Influence of avatar appearance on presence in social VR. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 233–234, Mar. 2017. doi: 10.1109/3DUI.2017.7893357

[9] J. Herder, N. Brettschneider, J. de Mooij, and B. Ryskeldiev. Avatars for co-located collaborations in hmd-based virtual environments. In *IEEE VR 2019, 26th IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 968 – 969. Osaka, Mar. 2019. doi: 10.1109/VR.2019.8798132

[10] J. Lacoche, N. Pallamin, T. Boggini, and J. Royan. Collaborators Awareness for User Cohabitation in Co-located Collaborative Virtual Environments. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, pp. 15:1–15:9. ACM, New York, NY, USA, 2017. doi: 10.1145/3139131.3139142

[11] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The Effect of Avatar Realism in Immersive Social Virtual Realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, pp. 39:1–39:10. ACM, New York, NY, USA, 2017. doi: 10.1145/3139131.3139156

[12] M. Lombard, T. B. Ditton, and L. Weinstein. Measuring Presence: The Temple Presence Inventory. In *Proceedings of the 12th Annual International Workshop on Presence*. Los Angeles, California, USA, Nov. 2009.

[13] J. L. Lugrin, M. Wiedemann, D. Bieberstein, and M. E. Latoschik. Influence of avatar realism on stressful situation in VR. In *2015 IEEE Virtual Reality (VR)*, pp. 227–228, Mar. 2015. doi: 10.1109/VR.2015.7223378

[14] Y. Pan and A. Steed. The impact of self-avatars on trust and collaboration in shared virtual environments. *PLoS ONE*, 12(12), Dec. 2017. doi: 10.1371/journal.pone.0189078

[15] B. Rasch, M. Friese, W. J. Hofmann, and E. Naumann. *Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler*. Springer-Lehrbuch. Springer-Verlag, Berlin Heidelberg, 3 ed., 2010.

[16] D. Roth, J. L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 277–278, Mar. 2016. doi: 10.1109/VR.2016.7504761

[17] M. Russo, R. De Luca, A. Naro, F. Sciarrone, B. Aragona, G. Silvestri, A. Manuli, A. Bramanti, C. Casella, P. Bramanti, and R. Salvatore Calabrò. Does body shadow improve the efficacy of virtual reality-based training with BTS NIRVANA?: A pilot study. *Medicine*, 96:e8096, Sept. 2017. doi: 10.1097/MD.0000000000008096

[18] J. S. Casanueva and E. Blake. The Effects of Avatars on Co-presence in a Collaborative Virtual Environment. In *Conference of the South African Institute of Computer Scientists and Information Technologists*. Pretoria, South Africa, Jan. 2001.

[19] H. Salzmann, J. Jacobs, and B. Froehlich. Collaborative Interaction in Co-Located Two-User Scenarios. In *EGVE/ICAT/EuroVR*, pp. 85–92, Jan. 2009. doi: 10.2312/EGVE/JVRC09/085-092

[20] M. Schrepp, A. Hinderks, and J. Thomaschewski. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, and A. Marcus, eds., *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, vol. 8517, pp. 383–392. Springer International Publishing, Cham, 2014. doi: 10.1007/978-3-319-07668-3_37

[21] M. Slater, A. Sadagic, M. Usoh, and R. Schroeder. Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. *Presence: Teleoperators and Virtual Environments*, 9(1):37–51, Feb. 2000. doi: 10.1162/105474600566600

[22] W. Steptoe, O. Oyekoya, A. Murgia, R. Wolff, J. Rae, E. Guimaraes, D. J. Roberts, and A. Steed. Tracking for avatar eye-gaze control during object-focused multiparty interaction in immersive collaborative virtual environments. In *Proceedings of IEEE Virtual Reality*, pp. 83–90. IEEE, Lafayette, Louisiana USA, Mar. 2009.

[23] K. Tcha-Tokey, O. Christmann, E. Loup-Escande, and S. Richir. Proposition and Validation of a Questionnaire to Measure the User Experience in Immersive Virtual Environments. *The International Journal of Virtual Reality*, 16(1):33–48, 2016.

[24] C. Wienrich, K. Schindler, N. Döllinger, S. Kock, and O. Traupe. Social Presence and Cooperation in Large-Scale Multi-User Virtual Reality – The Relevance of Social Interdependence for Location-Based Environments. In *IEEE VR 2018, the 25th IEEE Conference on Virtual Reality and 3D User Interfaces*. Reutlingen, Mar. 2018.