Effects on Co-Presence of a Virtual Human: A Comparison of Display and Interaction Types

Daehwan Kim¹ and Dongsik Jo²,*

¹ Content Research Division, ETRI, Daejeon 34129, Korea; daehwank@etri.re.kr
² School of IT Convergence, University of Ulsan, Ulsan 44610, Korea
* Correspondence: dongsikjo@ulsan.ac.kr; Tel.: +82-52-259-1647

Abstract: Recently, artificial intelligence (AI)-enabled virtual humans have been widely used in various fields in our everyday lives, such as for museum exhibitions and as information guides. Given the continued technological innovations in extended reality (XR), immersive display devices and interaction methods are evolving to provide a feeling of togetherness with a virtual human, termed co-presence. With regard to such technical developments, one main concern is how to improve the experience through the sense of co-presence as felt by participants. However, virtual human systems still have limited guidelines on effective methods, and there is a lack of research on how to visualize and interact with virtual humans. In this paper, we report a novel method to support a strong sense of co-presence with a virtual human, and we investigated the effects on co-presence with a comparison of display and interaction types. We conducted the experiment according to a specified scenario between the participant and the virtual human, and our experimental study showed that subjects who participated in an immersive 3D display with non-verbal interaction felt the greatest co-presence. Our results are expected to provide guidelines on how to focus on constructing AI-based interactive virtual humans.

Keywords: virtual human; co-presence; interaction; extended reality (XR)

1. Introduction

In the days ahead, the influence of virtual reality (VR) and augmented reality (AR), commonly referred to as extended reality (XR) in brief, is expected to grow rapidly. XR has received significant attention as a key technology for education, training, gaming, advertising, and shopping given its ability to add effective information for our everyday lives [1,2]. This technology can create various situations with useful datasets and interactions to improve a user’s experience [3]. In particular, it is actively used for situational training in place of actual people [4]. More recently, virtual humans—computer-generated characters—in XR were developed to act as guides to include real-world meaningful information combined with artificial intelligence (AI) technology. For instance, interactive virtual humans can empower users to collaborate in three-dimensional immersive tele-conferencing, during which one can communicate with a teleported remote other as if present [5,6]. Additionally, after reconstructing a virtual human through a captured real human, it can be used for the purpose of receiving responses with a participant, such as a virtual human who answers questions [7]. Moreover, with the development of collaborative virtual environments (CVEs), online games, and the increasing popularity of metaverses, virtual humans are becoming even more important, and this technology allows participants to interact with virtual humans as if they are actually present in the same room [8,9]. Despite such provisions, virtual human systems still have limited guidelines on effective methods, and there is a lack of research on how to visualize and interact with virtual humans. In particular, it is necessary to increase the sense of co-presence, i.e., the feeling of being together, via effective visualization and interaction methods with perceived virtual humans in XR environments.
In this line of thinking, herein, we propose a method to increase and improve the co-presence of virtual humans that starts with a comparison of different visual and interaction types. For example, imagine a situation in which a user is wearing a helmet-type immersive device; we need to provide a suitable interaction method, such as dialogue and gestures, for greater co-presence. It would be best to provide all cases of interactions for the XR virtual human; in some cases, due to given device setup environments and computing and development resources, we will need to know which interaction method is more effective. Our paper especially focuses on display and interaction elements for co-presence (e.g., 2D vs. 3D, verbal vs. non-verbal). We note that the main XR technologies consist of visualization, interaction, and simulation. Visualization and interaction are related to the setup of the XR virtual human within these technologies [10]. Thus, for an effective setup of a virtual human, we performed comparative experiments on different display and interaction types using a virtual human, and our experimental study with subjects participating in an immersive 3D display using a head-mounted display (HMD) showed that gesture-based non-verbal interaction is most effective. Additionally, for a 2D XR environmental situation, it was found that dialogue-based verbal interaction is more important.

Figure 1 shows typical interaction situations with a virtual human. The participant labeled “Real Human” in Figure 1 can communicate and interact via inputs such as voice and gestures with the perceived virtual human in the given display system, and the virtual human can provide appropriate response feedback such as facial expressions, answers, and interactive animation behaviors. In this case, a sensing device consisting of an integrated depth camera is mainly used to recognize the participant’s voice and gestures. Note that the depth camera has pixel values that correspond to the distance, allowing the extraction and estimation of the human’s body skeleton for gesture recognition; it also contains controllers for voice interactions [11]. In this case, we need to know which combinations can increase co-presence. This is the focus of this paper.

**Figure 1.** Display and interaction types with the virtual human and an example of a user’s interaction with a virtual human: our figure presents a scene in which a user’s gesture-based non-verbal interaction is given to the virtual human in an immersive 3D XR environment using a head-mounted display.

The remainder of our paper is organized as follows. Section 2 reviews works related to our research. Section 3 provides a system overview, and Section 4 describes the detailed implementation. Section 5 provides the experimental setup and procedure for evaluating co-presence in the system configuration. Section 6 reports the main results and provides
Section 7 summarizes the paper and concludes with directions for future works.

2. Related Works

2.1. Interactive Virtual Human

Life-sized virtual humans made to resemble actual humans have been widely used as an important media technology in our everyday life [12]. A virtual human can use language in an interactive way with appropriate gestures and can show emotional reactions to verbal and non-verbal stimuli [13]. Additionally, the virtual human can provide face-to-face communication and interaction with a remote teleport avatar [14]. A useful example with virtual humans is to apply them in training scenarios such as interviews and medical team training [15,16]. Another notable example using interactive virtual humans is museum application scenarios [17]. When a visitor enters the exhibit space, the virtual human appears on a stage in the museum room and can explain key necessary information, in the same way as a museum guide. Then, the visitor can ask questions and obtain answers from the virtual human. More recently, many researchers have been concerned with finding an intuitive way for virtual humans to control physical objects in augmented reality (AR) combining the real world and virtual humans [18,19]. It remains to be determined how these research results should be configured to achieve the highest level of co-presence with the virtual human to enhance the effect.

2.2. Display Types for the Virtual Human

In a few more recent results, many researchers have introduced XR displays to visualize a life-sized virtual human, such as head-worn displays, multi-projection for immersive environments, and auto-stereoscopic 3D displays [20,21]. As a future teleconference service scenario, multiple depth cameras and a see-through head-mounted display (named Hologportation) allowing precise 3D virtual human models to be transmitted to a remote site in real time were introduced by Escolano et al. [22]. Shin and Jo suggested a mixed-reality human (MRH) system in which the virtual human is combined with physicality as the real part [23]. According to recent requirements, Hologram showcased a holographic virtual stage and a holographic virtual character to make family life more convenient and the environment more intelligent [24]. Our work was designed to further improve upon pioneering works by comparing different types of displays in relation to the effects of interaction with a virtual human.

2.3. Interaction Types for the Virtual Human

A virtual human is an output form of computer systems that strive to engage with participants through natural language interfaces (or verbal interaction) and non-verbal interaction, which can also be performed with facial expressions and gestures [25,26], and the virtual human provides a response to the real person who interacts with it. Our work investigated such issues where it is desirable that the virtual human mimics the behavior of a real person.

Image-based full-body capture technology refers to three-dimensional reconstruction of the appearance of the body, hands, and face from image data, and this can also capture human movements and changes. With the recent development of deep learning technology and the rapid rise of XR, full-body capture technology can easily and quickly be applied to dynamic virtual humans [27]. In particular, with the development of deep neural network (DNN) algorithms, it has reached the level of estimating three-dimensional body poses and shapes for interaction [28,29]. However, there are still further research topics to be investigated for virtual human interaction, such as effective motion and feedback. In this paper, our focus is on finding an effective configuration among the various interaction types in terms of co-presence.
2.4. Co-Presence for the Virtual Human

There have been a few previous attempts to evaluate co-presence levels when interacting with a virtual human in various XR environments. As a representative research result, Mathis et al. explored the perception of virtual humans in virtual environments, and they presented the impact of virtual human’s fidelity with respect to interaction [30]. According to another study, Slater et al. conducted an experiment in which participants responded appropriately to a negative or positive virtual audience, and the result showed virtual humans are more effective when perceived as humans [31]. An experiment by Zanbaka et al. was reported in which participants were inhibited by the co-presence (being together) of virtual humans in performing a task [32]. Additionally, Jo et al. presented the effects of the type of virtual human and the background representation on the participants’ co-presence with respect to the design of future virtual humans [33]. Then, co-presence is defined as feeling that the other participants in the virtual environment exist [34].

To measure the user’s feeling of being together, called co-presence, with the virtual human, the previous studies usually conducted surveys via questionnaire [34,35]. More recently, physiological measurements of parameters such as heart rate and skin conductance were used to investigate the effects on co-presence under different XR conditions [36,37]. However, most previous works have focused on the shape of the virtual human, such as gaze direction and animation qualities, and no comprehensive work has been done in terms of the effect of different display and interaction types. Thus, our work investigated such issues regarding visual and interaction components.

3. System Overview

Table 1 shows several examples of test conditions for our experiment to determine their effects on co-presence with the virtual human (e.g., various types of display, such as a head-mounted display, a large display to visualize a life-scaled virtual human, and multiple projectors, as well as verbal or non-verbal interaction methods). Usually, participants are in a VR/AR system with a virtual human experience stereoscopic visualization; more recently, large 2D displays have been used to overcome the inconvenience of wearing a head-mounted display (HMD) [23]. As another issue, when interacting with a virtual human, just as people communicate with each other in everyday life, participants can use their voice as well as gesture and facial expressions. Interaction methods are divided into conversation-based verbal methods and non-verbal characteristics that do not depend on dialogue. In our system, the virtual human was designed to provide verbal communication or non-verbal interaction in the given display systems, such as the 2D mono or 3D stereoscopic types, and the virtual human could automatically give the participant appropriate response feedback, such as facial expressions, answers, and interactive animation behaviors, according to the participant’s input.

Table 1. Candidate test conditions in our experiment in terms of display and interaction types.

| Test Conditions     | Methods                  | Characteristics |
|---------------------|--------------------------|-----------------|
| Display Types       | Head-mounted display     | 3D              |
|                     | Life-size supported TV   | 2D              |
|                     | Multi-projection         | 3D              |
|                     | Holographic Image        | 3D              |
| Interaction Types   | Voice for Conversation   | Verbal          |
|                     | Gesture                  | Non-verbal      |
|                     | Facial Expression        | Non-verbal      |

For example, Figure 2 shows our interaction handling process between the user and the virtual human through the recognition of various verbal and non-verbal inputs to establish the virtual human’s response [19]. For example, to determine the user’s gesture, the user’s specific joint position information is extracted from the depth camera and the
user’s joint position information is then calculated. Subsequently, the system recognizes whether the user performed a gesture at a specific point. At this time, if the score of the highest motion among the pre-trained candidate motions is higher than a pre-defined threshold \(T\), we determine that the user performed the corresponding gesture operation. In addition, cases involving other motions or an absence of motions are set to “None”, and a case with a score lower than \(T\) is recognized as “None” as well.

[Table 1. Candidate test conditions in our experiment in terms of display and interaction types.]

| Test Conditions | Methods | Characteristics |
|-----------------|---------|----------------|
| Display Types   |         |                |
| Head-mounted display | 3D |                |
| Life-size supported TV | 2D |                |
| Multi-projection | 3D |                |
| Holographic Image | 3D |                |
| Interaction Types |         |                |
| Voice for Conversation | Verbal |                |
| Gesture | Non-verbal |                |
| Facial Expression | Non-verbal |                |

[Figure 2. Verbal and non-verbal interaction and response matching process of the virtual human.]

4. Implementation

To evaluate co-presence with virtual humans, first we developed an authoring toolkit to edit the virtual human’s responses to apply the user’s interactions [38]. The virtual human needs to generate an interaction response, and it should be possible to control the lip-sync motion with the voice to make conversational situations and gesture-based animation to create appropriate behaviors matching those of a real human. Thus, our authoring toolkit to provide the virtual human controller allows us to create the virtual human’s responses (e.g., voice-based verbal conversation, facial expressions, and gesture-based non-verbal feedback) and realize a context-aware virtual human for our everyday lives [39]. In other words, it provides a quick pipeline to build interactive virtual humans to support verbal and non-verbal situations (see Figure 3).

In test conditions for immersive 3D displays, the subject wore a head-mounted display (HTC VIVE [40]) to view and interact with the virtual human. On the other hand, we installed the virtual human on a large TV display (an 85-inch screen in portrait mode) for 2D situations to show the same life-sized virtual human that would be presented in a 3D immersive display. The motion datasets of the virtual human, such as the appearances of its facial expressions, its animated motions, and lip-synching with prerecorded sounds, were configured with resemblance to an actual person using Mixamo and Oculus LipSync in Unity3D [41,42].
were extracted from the depth camera, and each intended gesture was predicted based on adjustments made so that sound of the same volume and sound quality was output from the same speaker. To find the user’s body motion, the process of obtaining skeleton data was configured with resemblance to an actual person using Mixamo and Oculus LipSync Toolkit [42].

In test conditions for immersive 3D displays, the subject wore a head-mounted display. Figure 4 shows examples of a virtual human’s lip movements when providing verbal responses. With audio input streams with the user’s voice from the installed multi-array microphone, we predicted the lip movements and facial expressions that correspond to a particular voice and used a mouthpiece to animate the virtual human [42].

In real-time user interaction situations with verbal and non-verbal inputs, the virtual human can recognize users’ situations, presenting an optimal response to users by context awareness.

To express the whole-body motion of the virtual human, we selected candidates mapped to the suggested authoring toolkit that dynamically respond to changes in the actual human’s motion configuration. In this case, 32 specific joint positions of the user were extracted from the depth camera, and each intended gesture was predicted based on the user’s joint position information. Joint information was used to recognize whether the user performed a gesture at a specific point in time. Additionally, to express more sufficient virtual human motion, the virtual human’s gaze and eyeball orientation were automatically transformed by recognizing the user’s location and the locations of sounds in the surrounding environment. Figure 5 shows examples of the motions of the virtual human motion, the virtual human’s gaze and eyeball orientation were automatically transformed.

Fuzzy logic is a form of many-valued logic in which the truth values of variables can be any real number between 0 and 1, inclusive. We used a depth camera and a multi-array microphone to capture the user’s voice and body gestures for the same sound conditions, with adjustments made so that sound of the same volume and sound quality was output from the same speaker. To find the user’s body motion, the process of obtaining skeleton information was carried out based on specific joints of the user [44].

Figure 4 shows examples of the virtual human’s lip movements when providing verbal responses. With audio input streams with the user’s voice from the installed multi-array microphone, we predicted the lip movements and facial expressions that correspond to a particular voice and used a mouthpiece to animate the virtual human [42].
automatically transformed by recognizing the user’s location and the locations of sounds in the surrounding environment. Figure 5 shows examples of the motions of the virtual human that we set up as candidates for our experiment. The results of the virtual human responses were expressed in a life-sized form through the given display devices.

![Figure 5. Examples of virtual human motion that we set up as candidates: the virtual human’s motion types consisted of idle, applause, shy, surprised, and so on.](image)

Figure 5 shows examples of the motions of the virtual human that we set up as candidates for our experiment. The results of the virtual human responses were expressed in a life-sized form through the given display devices.

We leave it as future work to provide and match a greater range of motions from daily life.

Figure 6 shows an example of the interaction results between the user and the virtual human. This figure shows a scene that provides responses to the user’s non-verbal inputs, such as gestures and emotions. As noted earlier, we used the toolkit to match the interactive animation of the virtual human according to the user’s input with inference via fuzzy logic. We leave it as future work to provide and match a greater range of motions from daily life. During the experiment, the virtual human was visualized on a personal computer running a 64-bit version of Windows 10.

![Figure 6. Interaction results between a user and a virtual human: (a) The virtual human represented in a large 2D TV. (b) According to gesture recognition, the virtual human can express the specific motion (e.g., applause animation) responding to the user. This motion was the result of optimal responses related to the virtual human’s behavior by the participant’s actions and the recognized situations despite showing the mirroring gesture.](image)
5. Experiment to Determine Effects on Co-Presence

We now describe an experiment that examined how people respond to two different conditions, labeled here as display and interaction types. The experiment was designed as a two-factor (two levels each) between-subject measurement. The first factor was display type (a large 2D TV vs. a 3D immersive HMD), and the second factor was the interaction type (verbal vs. non-verbal). Thus, we examined four different configurations, as shown in Table 2. Note that our experiment excluded duplicate characteristics, as indicated in Table 1, and 3D and non-verbal situations were considered and selected.

| Test Conditions | Display Types          | Interaction Types       |
|-----------------|------------------------|-------------------------|
| 3D-verbal       | Head-mounted display (3D) | Voice for Conversation (Verbal) |
| 2D-verbal       | Life-size supported TV (2D) | Voice for Conversation (Verbal) |
| 3D-non-verbal   | Head-mounted display (3D) | Gesture (Non-verbal)    |
| 2D-non-verbal   | Life-size supported TV (2D) | Gesture (Non-verbal)    |

For example, in the 3D-verbal test conditions, the subject wore a head-mounted display (HTC Vive) and was allowed to talk only with the virtual human. On the other hand, the 2D-non-verbal situation was designed to allow interaction gestures in front of a life-sized display that visualized the virtual human at the same size used in the 3D condition.

In the experiment, to assess the level of user-felt co-presence, the subjects attempted to have a conversation with the interactive virtual human. We adopted the simulated method of a greeting scenario to engage in typical conversations or gestures with two people, where a student in a class attends lessons at a university, as proposed by Shin and Jo [45]. Then, because most of the participants in the experiment were college students, the participants played the role of the student, and the virtual human was expressed as a teacher. Verbal questions asked or non-verbal gestures performed by the participant and directed toward the virtual human were presented as given candidates in the common conversation, and the virtual human’s verbal answers and non-verbal gestures were inferred via fuzzy logic using the previously introduced toolkit. Each participant was allowed to interact with the virtual human (See Figure 7 for the detailed process). In our experiment, we utilized 50 verbal sentences and seven non-verbal motions.
Forty paid subjects (25 men and 15 women) with a mean age of 27.8 years participated in the between-subject experiment. The participants in the experiment were divided into four groups of ten each, and the groups were equally divided based on the prior experience of virtual reality (VR). The experiment was designed as a two-factor experiment (two display types × two interaction types), and every participant experienced only one condition to avoid any learning effects in various conditions. Note that a within-subject experimental design was not used due to the learning effect. Before carrying out the task for the given treatment, the participants were initially briefed on the overall purpose of the experiment for approximately five minutes, with the virtual teacher leading the conversation or providing a greeting gesture. Each participant took part in the interaction form of greeting for three minutes with the virtual teacher. Our task simulated the virtual human’s interaction situations, which allows participants to give and receive responses to the specific questions and behaviors [46]. Participants can receive verbal or non-verbal feedback from the virtual human such as greetings at the start of a new semester. Our system also provided participants’ response candidates for the simulated question and the given gesture of the virtual human, which gives related interaction to the subject. If the participant did not act or say anything, the virtual human reacted by speaking, making gestures, and engaging in movements according to the control of the administrator behind the curtain, referred to here as Wizard-of-Oz testing [23]. Upon completion of the task, the participants’ sense of co-presence was measured via a survey. Here, we defined co-presence as how much a subject perceived someone to be in the same space [34]. Additionally, we investigated immersive tendencies scores (ITQs) of the virtual human in given conditions using Witmer and Singer’s immersive tendencies questionnaire [35]. The following questions were posed: (1) “When interacting with the virtual human, do you become as involved as you would in the actual situation?” and (2) “Did you feel as if you were present with the virtual human in the same space?” The subjects completed a quantitative questionnaire to assess the degrees of immersion and co-presence on a seven-point Likert scale. We also collected qualitative feedback.

6. Results and Discussion

First, immersion was judged in situations with an actual person to establish the “ground truth”. We also included a baseline case and projected a video of the actual person onto the display system before starting the experiment in each condition. We compared the immersion scores generated by the different conditions through a one-way ANOVA. The results did not reveal statistically significant main effects regarding the level of immersion in the four conditions, $F(2,15) = 32.89, p > 0.05$ (See Figure 8). On the other hand, interaction with the virtual human was found to have positive effects on immersion (in all cases over average four-point scores).

![Error Bars: 95% Confidence Intervals](Figure 8. Level of immersion among the four test conditions.)
Second, two-sample t-tests between non-verbal and verbal conditions (i.e., 3D–non-verbal vs. 3D–verbal or 2D–non-verbal vs. 2D–verbal) revealed significant main effects with \( p < 0.05 \) regarding the level of co-presence. A pairwise comparison showed a main effect in the statistical analysis (3D–non-verbal and 3D–verbal, \( t(16.35) = 7.48, p < 0.05 \); 2D–non-verbal and 2D–verbal, \( t(17.967) = -8.05 \)). Note that a t-test can be used to determine if the means of two sets of data are significantly different from each other [47]. When we plotted the distribution of the values of the sample, it was possible to infer that a normal distribution was established on a plot. Then, we found that there was a significant difference from our result in an unequal variance situation where more reliable data can be estimated. Additionally, we used the Wilcoxon signed-rank test for a non-parametric statistical hypothesis; this method is used when the distribution is not normal. The same result was obtained. In our results, the level of co-presence was clearly different for each configuration in the non-verbal and verbal interactions (See Figure 9). Participants who used the immersive 3D display reported that non-verbal interaction was more effective; in the case of the 2D XR environmental situation, it was found that dialogue-based verbal interaction improved the user’s sense of being in the same space (co-presence). Through subjective answers by an additional questionnaire completed by the participants, we found that verbal interaction in 2D was more effective in terms of co-presence, as the participants feel more familiar with this setup given their possible experience with video conferencing systems. On the other hand, in the 3D immersive environment, the 3D interaction method (e.g., non-verbal gestures) scored higher than the verbal approach, and it is expected to be helpful in terms of co-presence. Thus, participants felt that verbal interaction was familiar as a form of interaction with the 2D environment, while they commented that gesture-based non-verbal interaction was more suitable for the 3D environment. Moreover, two-sample t-tests between the 2D and 3D conditions (i.e., 2D–verbal vs. 3D–verbal or 2D–non-verbal vs. 3D–non-verbal) indicated significant main effects with \( p < 0.05 \) as well. The pairwise comparison showed a main effect in the statistical analysis (2D–verbal and 3D–verbal, \( t(16.89) = 7.47, p < 0.05 \); 2D–non-verbal and 3D–non-verbal, \( t(17.789) = -8.06 \)). Then, the same statistical analysis method as already mentioned above was applied. The level of co-presence was significantly different for each configuration in 2D and 3D. As noted earlier, subjects in the immersive 3D display environment using a head-mounted display (HMD) reported that gesture-based non-verbal interaction was most effective. For the 2D environment, we found that dialogue-based verbal interaction was more important with respect to co-presence with the virtual human.

![Figure 9. Level of co-presence among the four test conditions. In this figure, we compared non-verbal and verbal using a pairwise comparison. A plus sign denotes the average.](image-url)
For a more accurate analysis, we also applied a qualitative analysis with feedback from the participants [31]. It was conducted in the form of three short questions and answers. The questions are as follows: “What were your criteria regarding immersion?”, “What were your criteria regarding co-presence?”, and “What do you think needs to be done to achieve higher co-presence with the virtual human?” In the results, there was a consensus in that the participants reported that the overall 3D environment appeared to be high in immersion, but a major difference was noted. This is identical to the result found in the statistics. Interestingly, there were several answers to the question of focusing on co-presence criteria. A few participants mentioned that the 3D interaction method was familiar in a 3D environment (e.g., gesture-based), whereas the 2D interaction method, i.e., verbal-based interaction, helped with co-presence in the 2D environment. One participant reported that it would improve the feeling of co-presence to provide possible interactions all at once (e.g., multimodal interactions) rather than a single input. In order to obtain the result of this situation, we will leave it as a future work to determine this improvement, if any, when interacting with a virtual human. Additionally, some participants mentioned limitations of the VR equipment preventing an increase to higher co-presence (e.g., the lightweight headsets, precise interaction, and an immersive display). Thus, we plan to conduct an experiment in the future using an immersive projection display space such as CAVE with a wide field of view (FOV) [48].

With our results, we presented a preliminary implementation of interaction with a virtual human, partly validating the effectiveness of a future virtual human system. Moreover, we were able to present a setup guideline for the virtual human. There are still many aspects that need to be improved to provide full guidelines on how to increase co-presence with the virtual human. Additionally, to increase internal reliability of our study, we hope to provide the same environmental configuration in 2D and 3D. In our situation, the experiment was conducted by implementing the same 2D surrounding environment as far as possible with 3D modeling. We will need to overcome the hardware experimental factors for the weight of the head-mounted display in 3D to improve threats to external validity. Furthermore, it will be necessary to evaluate and investigate co-presence by presenting the various situations of interaction with a real person, such as a job interview.

7. Conclusions and Future Works

Recent advances in virtual human technologies will provide more helpful information to assist in real-world tasks and to increase the level of intelligence in our lives. Herein, we presented the effects of display and interaction types on the participants’ feeling of co-presence—the feeling of being together with the perceived virtual human—for the design of virtual human interaction systems. We focused on the co-presence of the virtual human and used an approach focusing on the perception from the user’s intentions, such as when they used their voice and gestures. In our study, participants who used an immersive 3D display with a head-mounted display (HMD) reported that gesture-based non-verbal interaction was most effective. On the other hand, in the case of the 2D XR environmental situation used here, it was found that dialogue-based verbal interaction improved the user’s sense of sharing and being in the same space. These findings have implications for the design of more effective virtual human systems that offer a high sense of co-presence.

In future works, we will continue to explore the effects of other related factors to improve co-presence with the virtual human. We also hope to apply interaction response methods to generate the behavior of the virtual human using AI learning algorithms to improve the quality of interaction and employ physiological measures to assess co-presence quantitatively in various situations.
Author Contributions: D.K. performed the prototype implementation and co-presence experiments; D.J. designed the study, performed statistical analysis of the results, and contributed to the writing of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the 2021 Research Fund of the University of Ulsan.

Institutional Review Board Statement: The participant’s personal identification information used in the study did not include personal information. Ethical review and approval were not required for the study.

Informed Consent Statement: Written informed consent has been obtained from the interview and user study participants to publish this paper.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jo, D.; Kim, K.-H.; Kim, G.J. Spacetime: Adaptive control of the teleported avatar for improved AR tele-conference experience. Comput. Animat. Virtual Worlds 2015, 26, 259–269. [CrossRef]
2. Xue, H.; Sharma, P.; Wild, F. User satisfaction in augmented reality-based training using Microsoft Hololens. Computers 2019, 8, 9. [CrossRef]
3. Shin, K.-S.; Kim, H.; Lee, J.; Jo, D. Exploring the effects of scale and color difference on users’ perception for everyday mixed reality (MR) experience: Toward comparative analysis using MR devices. Electronics 2020, 9, 1623. [CrossRef]
4. Robb, A.; Kopper, R.; Ambani, R.; Qayyum, F.; Lind, D.; Su, L.-M.; Lok, B. Leveraging virtual humans to effectively prepare learners for stressful interpersonal experiences. IEEE Trans. Vis. Comput. Graph. 2013, 19, 662–670. [CrossRef] [PubMed]
5. Pluss, C.; Ranieri, N.; Bazin, J.-C.; Martin, T.; Laffont, P.-Y.; Popa, T.; Gross, M. An immersive bidirectional system for life-size 3D communication. In Proceedings of the 29th International Conference on Computer Animation and Social Agents, Geneva, Switzerland, 23–25 May 2016.
6. Susumu, T. Telexistence: Enabling humans to be virtually ubiquitous. IEEE Comput. Graph. Appl. 2016, 36, 8–14.
7. Kang, S.; Krum, D.; Khooshabeh, P.; Phan, T.; Kevin Chang, C.; Amir, O.; Lin, R. Social Influence of Humor in Virtual Human Counselor’s Self-Disclosure. J. Comput. Animat. Virtual Worlds 2017, 28, e1763. [CrossRef]
8. Casanueva, J.; Blake, E. The effects of avatars on co-presence in a collaborative virtual environment. In Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists, Pretoria, South Africa, 25–28 September 2001.
9. Beck, S.; Kunert, A.; Kulik, A.; Froehlich, B. Immersive group-to-group telepresence. IEEE Trans. Vis. Comput. Graph. 2013, 19, 616–625. [CrossRef] [PubMed]
10. Johnson, S.; Orban, D.; Runesha, H.; Meng, L.; Juhnke, B.; Erdman, A.; Samsel, F.; Keefe, D. Bento Box: An Interactive and Zoomable Small Multiples Technique for Visualizing 4D Simulation Ensembles in Virtual Reality. Front. Robot. AI 2019, 6, 61. [CrossRef]
11. Carraro, M.; Munaro, M.; Menegatti, E. Skeleton estimation and tracking by means of depth camera fusion form depth camera networks. Robot. Auton. Syst. 2018, 110, 151–159. [CrossRef]
12. Peja, T.; Kantor, J.; Benko, H.; Ofek, E.; Wilson, A. Room2Room: Enabling life-size telepresence in a projected augmented reality environment. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing, San Francisco, CA, USA, 27 February–2 March 2016.
13. Volante, M.; Babu, S.-V.; Chaturvedi, H.; Newsome, N.; Ebrahimi, E.; Roy, T.; Daily, S.-B.; Basolo, T. Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. IEEE Trans. Vis. Comput. Graph. 2016, 22, 1326–1335. [CrossRef]
14. Maimone, A.; Fuchs, H. A First Look at a Telepresence System with Room-Sized Real-Time 3D Capture and Large Tracked Display Wall. In Proceedings of the 21st International Conference on Artificial Reality and Telexistence (ICAT), Osaka, Japan, 28–30 November 2011.
15. Kwon, J.-H.; Powell, J.; Chalmers, A. How level of realism influences anxiety in virtual reality environments for a job interview. Int. J. Hum.-Comput. Stud. 2013, 71, 978–987. [CrossRef]
16. Robb, A.; Cordar, A.; Lampotang, S.; White, C.; Wendling, A.; Lok, B. Teaming up with virtual humans: How other people change our perceptions of and behavior with virtual teammates. IEEE Trans. Vis. Comput. Graph. 2015, 21, 511–519. [CrossRef] [PubMed]
17. Duguleana, M.; Briciu, V.-A.; Duduman, I.-A.; Machidon, O. A virtual assistant for natural interactions in museums. Sustainability 2020, 12, 6958. [CrossRef]
18. Yoon, L.; Yang, D.; Chung, C.; Lee, S.-H. A mixed reality telepresence system for dissimilar spaces using full-body avatar. In Proceedings of the SIGGRAPH ASIA XR, Online, 10–13 December 2020.
19. Kim, D.; Jo, D. Exploring the effects of gesture interaction on co-presence of a virtual human in a hologram-like system. *J. Korea Inst. Inf. Commun. Eng.* 2020, 24, 1390–1393.

20. Maimone, A.; Yang, X.; Dierk, N.; State, A.; Dou, M.; Fuchs, H. General-purpose telepresence with head-worn optical see-through displays and projector-based lighting. In Proceedings of the IEEE Virtual Reality, Orlando, FL, USA, 16–20 March 2013.

21. Jones, A.; Lang, M.; Fyffe, G.; Yu, X.; Busch, J.; Medowall, I.; Debevec, P. Achieving eye contact in a one-to-many 3D video teleconferencing system. *ACM Trans. Graph.* 2009, 28, 64. [CrossRef]

22. Orts, S.; Rhemann, C.; Fanello, S.; Kim, D. Holoportation: Virtual 3D teleportation in real-time. In Proceedings of the 29th ACM User Interface Software and Technology Symposium (UIST), Tokyo, Japan, 16–19 October 2016.

23. Shin, K.-S.; Kim, H.; Jo, D. Exploring the effects of the virtual human and physicality on co-presence and emotional response. *J. Korea Soc. Comput. Inf. 2019, 24, 67–71. [CrossRef]

24. Gao, H.; Xu, F.; Liu, J.; Dai, Z.; Zhou, W.; Li, S.; Yu, Y. Holographic three-dimensional virtual reality and augmented reality display based on 4K-spatial light modulators. *Appl. Sci. 2019, 9, 1182. [CrossRef]

25. Kim, K.; Schubert, R.; Hochreiter, J.; Bruder, G.; Welch, G. Blowing in the wind: Increasing social presence with a virtual human via environmental airflow interaction in mixed reality. *Elsevier Comput. Graph. (CAG) 2019, 83, 23–32. [CrossRef]

26. Wang, I.; Ruiz, J. Examining the use of nonverbal communication in virtual agents. *Int. J. Hum.-Comput. Interact. 2021, 37, 1–26. [CrossRef]

27. Zhao, Y. Research on virtual human animation based on motion capture data. In Proceedings of the International Conference on Data Processing Techniques and Applications for Cyber-Physical Systems, Guangxi, China, 11–12 December 2020.

28. Mathis, A.; Schneider, S.; Lauer, J.; Mathis, M. A primer on motion capture with deep learning principles, pitfalls, and perspectives. *Neuron 2020, 108, 44–65. [CrossRef]

29. Chentanez, N.; Muller, M.; Macklin, M.; Makoviychuk, V.; Jeschke, S. Physics-based motion capture imitation with deep reinforcement learning. In Proceedings of the 13th Annual ACM SIGGRAPH Conference on Motion, Interaction and Games (MIG), Limassol, Cyprus, 8–10 November 2018.

30. Mathis, F.; Vaniea, K.; Khamis, M. Observing virtual avatars: The impact of avatars’ fidelity on identifying interactions. In Proceedings of the Mindtrek 2021, Tampere, Finland, 1–3 June 2021; pp. 154–164. [CrossRef]

31. Slater, M.; Pertaab, D.; Steed, A. Public speaking in virtual reality: Facing an audience of avatars. *IEEE Comput. Graph. Appl. 1999, 19, 6–9. [CrossRef]

32. Zanbaka, C.; Ullinski, A.; Goolkasian, P.; Hodges, L.F. Effects of virtual human presence on task performance. In Proceedings of the 14th International Conference on Artificial Reality and Telexistence (ICAT), Seoul, Korea, 30 November 2004.

33. Jo, D.; Kim, K.-H.; Kim, G.J. Effects of avatar and background types on users’ co-presence and trust for mixed reality-based teleconference systems. In Proceedings of the 30th Conference on Computer Animation and Social Agents (CASA), Seoul, Korea, 22–24 May 2017.

34. Slater, M. Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence Teleoperators Virtual Environ. 1999, 8, 560–565. [CrossRef]

35. Witmer, B.G.; Singer, M.J. Measuring presence in virtual environments: A presence questionnaire. *Presence Teleoperators Virtual Environ. 1998, 7, 225–240. [CrossRef]

36. Meehan, M.; Razzaque, S.; Whitton, M.-C.; Brooks, F. Effect of latency on presence in stressful virtual environments. In Proceedings of the IEEE Virtual Reality, Orlando, FL, USA, 16–20 March 2013.

37. Zanbaka, C.; Ullinski, A.; Goolkasian, P.; Hodges, L.F. Effects of virtual human presence on task performance. In Proceedings of the 14th International Conference on Artificial Reality and Telexistence (ICAT), Seoul, Korea, 30 November 2004.

38. Chung, J.-H.; Jo, D. Authoring toolkit for interaction with a virtual human. In Proceedings of the Korea Information Processing Society Conference, Yeosu, Korea, 4–6 November 2021.

39. Latoschik, M.; Roth, D.; Gall, D.; Achenbach, J.; Waltemate, T.; Botsch, M. The effects of avatar realism in immersive social virtual realities. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology 2017, Gothenburg, Sweden, 8–10 November 2017.

40. Diederick, C.; Li, L.; Lappe, M. The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research. *I-Perception 2017, 8, 1–23. [CrossRef]

41. Yumak, Z.; Brink, B.; Egges, A. Autonomous Social Gaze Model for an Interactive Virtual Character in Real-Life Settings. *Comput. Animat. Virtual Worlds 2017, 28, e1757. [CrossRef]

42. Charalambous, C.; Yumak, Z.; Stappen, A. Audio-driven emotional speech animation for interactive virtual characters. *Comput. Animat. Virtual Worlds 2019, 30, e1892. [CrossRef]

43. Baijai, A.; Kushwah, V. Importance of fuzzy logic and application areas in engineering research. *Int. J. Recent Technol. Eng. 2019, 7, 1467–1471. [CrossRef]

44. Ma, C.; Zhang, S.; Wang, A.; Qi, Y.; Chen, G. Skeleton-based dynamic hand gesture recognition using an enhanced network with one-shot learning. *Appl. Sci. 2020, 10, 3680. [CrossRef]

45. Shin, E.; Jo, D. Mixed reality classroom based on multi-camera videos. In Proceedings of the Korea Institute of Information and Communication Engineering, Online, Korea, 30 October 2020; pp. 397–399.

46. Smith, M.; Ginger, E.; Wright, K.; Wright, M.; Taylor, J.; Humm, L.; Olsen, D.; Bell, M.; Fleming, M. Virtual reality job interview training in adults with autism spectrum disorder. *J. Autism Dev. Disord. 2014, 44, 2450–2463. [CrossRef]
47. Suarez, G.; Jung, S.; Lindeman, R. Evaluating virtual human role-players for the practice and development of leadership skills. *Front. Virtual Real.* 2021, 2, 658561. [CrossRef]

48. Cortes, C.; Argelaguet, F.; Marchand, E.; Lecuyer, A. Virtual shadows for real human in a CAVE: Influence on virtual embodiment and 3D interaction. In Proceedings of the 15th ACM Symposium on Applied Perception 2018, Vancouver, BC, Canada, 10–11 August 2018; pp. 1–8.