Multimodal Technologies and Interaction

MDPI

Article

Effect of Body Representation Level of an Avatar on Quality of AR-Based Remote Instruction

Tzu-Yang Wang [1,*], Yuji Sato [1], Mai Otsuki [2], Hideaki Kuzuoka [3] and Yusuke Suzuki [4]

[1] Department of Intelligent Interaction Technologies, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan; yuzi.march26since1996@gmail.com

[2] Human Augmentation Research Center, National Institute of Advanced Industrial Science and Technology, Kashiwa, Chiba 277-0882, Japan; mai.otsuki@aist.go.jp

[3] Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan; kuzuoka@cyber.t.u-tokyo.ac.jp

[4] OKI Electric Industry Co., Ltd., Warabi, Saitama 335-8510, Japan; suzuki543@oki.com

* Correspondence: st900278@gmail.com

Received: 26 November 2019; Accepted: 26 January 2020; Published: 5 February 2020

Abstract: In manufacturing, augmented reality (AR)-based remote instruction systems, which enable workers to receive instructions from an avatar, are widely used. In this study, we developed such a system and investigated the effect of the body representation level of the avatar on the quality of AR-based remote instruction. Drawing on the avatar designs of previous works, three different avatar designs ("Hand only", "Hand + Arm", and "Body"), representing three body representation levels, were created. In the experiment with a within-participant design, the avatar pointed at blocks sequentially and participants touched each block as soon as they identified it. The results of the experiment indicate that an AR-based remote instruction system with a "Body" avatar exhibits higher usability and can enable the participants to have a lower workload and higher efficiency.

Keywords: remote instruction; workload; efficiency; performance; preference; usability; augmented reality

1. Introduction

Among all types of collaboration, we focus on remote instruction, a typical type of collaboration in manufacturing. In remote instruction, an instructor at a remote location (i.e., away from the factory) gives instructions to a worker and the worker completes the task (such as connecting cables or picking up/putting down objects) based on the instructions.

At present, immersive 3D-based virtual reality (VR)/augmented reality (AR) technologies are widely used to support remote instruction, where users can collaborate as they are in the same virtual location, even when they are in different physical locations. Compared with remote instruction performed using 2D displays (e.g., tablet PCs and laptops), VR/AR-based remote instruction systems have the advantages of facilitating independent viewpoints, better spatial understanding, and realization of a shared workspace.

Body information of users (e.g., pertaining to the hand, head, and body) has recently been incorporated into VR/AR-based remote collaboration/instruction systems. Several studies have proved that representing body information improved the social presence of users and made remote collaboration more similar to face-to-face collaboration [1]. However, some other studies have highlighted the negative effects of body representation; for instance, users felt that their personal spaces were being invaded while collaborating with users with visualized arms [2]. Thus, to optimize the effects of body representation in an AR-based instruction system, it is desirable to select a suitable body representation level.

Instead of social factors (e.g., social presence), efficiency (how the system reduces the task completion time), performance (how the system improves the outcome of tasks), and workload (how the system reduces worker's workloads) have been considered more important in manufacturing. However, there has been little research focusing on the effects of body representation on these aspects relating to manufacturing.

Therefore, this research was aimed at investigating the effects of the body representation level on these aspects in an AR-based remote instruction system. An experiment was performed, in which participants were asked to interact with three types of avatars ("Hand only", "Hand + Arm", and "Body"), which represented three body representation levels (see Figure 1). To evaluate the effects of body representation level, the following five aspects related to manufacturing were considered: performance, efficiency, workload, usability, and preference. The contribution of this study was to explore a suitable body representation for AR-based remote instruction in manufacturing.

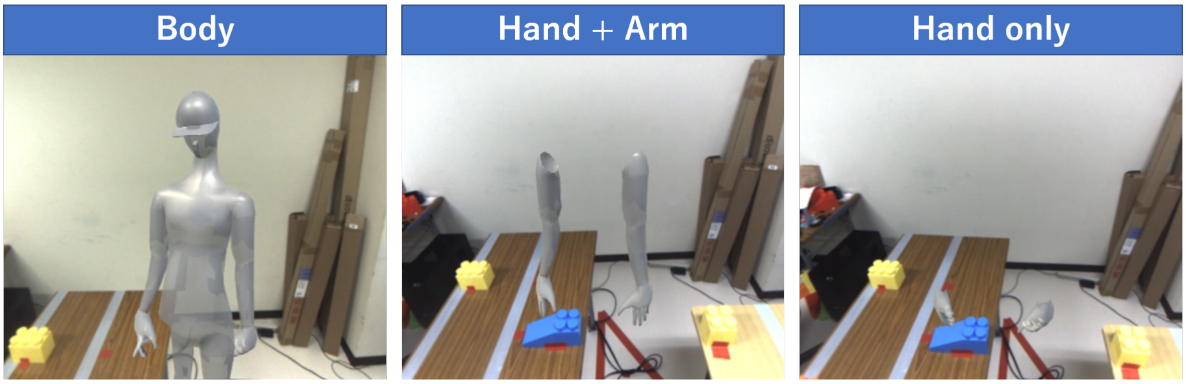

Figure 1. Three types of avatars representing different body representation levels.

2. Background

2.1. VR/AR-Based Remote Collaboration/Instruction

In the past two decades, several studies have focused on the use of immersive 3D-based VR/AR systems to improve remote collaboration and instruction. In contrast to 2D-based remote collaboration systems, which use 2D displays, such systems have several advantages and are now widely used.

First, one of the advantages is that it is easy for VR/AR to provide independent viewpoints for users. Shu and Flowers indicated that providing independent viewpoints for different participants could optimize parallel activities in a 3D CAD system [3]. Tait et al. also indicated that the provision of independent viewpoints led to faster task completion in remote collaboration [4]. By using a head-mounted display (HMD) and tracking system, users can freely move to set their viewpoint independently in such VR/AR-based systems [5,6].

Furthermore, VR/AR-based remote collaboration/instruction systems create immersive virtual environments which help to improve spatial understanding. Pausch et al. suggested that participants using HMDs had a better ability to locate targets in a virtual environment than when using a 2D desktop [7]. The use of an immersive virtual environment could help participants to maintain orientation in a given space [8] and provide spatial cues for navigating [9] and searching [10]. In addition, Huang indicated that the use of a 3D system could help users to understand the spatial relationships among objects [11].

Several researchers have also suggested that shared workspaces and attention play a key role in improving the efficiency of collaboration [12–16]. Wittaker suggested that, among all types of tasks in remote collaboration, a shared workspace had a significant effect on tasks requiring physical manipulation, such as moving and assembling objects [17]. AR-based remote collaboration/instruction systems provide better realization of shared workspace and shared attention. For example, in the AR instruction system of Higuchi et al., the working environment was shared with a remote instructor through a camera on an HMD; consequently, the remote instructor had a better understanding of the environment. Furthermore, the remote instructor's gaze was captured and shown in the worker's HMD; thus, the worker could better understand the aspects of the remote instructor's attention [18].

Another interesting research domain is the incorporation of a user's body information in VR/AR remote collaboration/instruction systems. The remote instruction system "HandsIn3D" of Huang et al. [11,19] used a 3D camera to capture both the hands of a helper and the worker space. Later, 3D meshes of the helper's hands were added into the meshes of the worker space, and the mixed meshes were shown in the worker's 2D display. In another study, HMDs were used to develop more immersive remote collaboration systems. The environments of local users were captured and reconstructed in a virtual environment, which was shown in the HMDs of remote users. In addition, the remote user's movements were captured and presented as an avatar to local users [5,20,21]. In contrast to these systems using HMD, Pejsa et al. used projected augmented reality techniques instead of HMDs to develop a life-sized co-present environment including the complete body of a remote participant for remote communication [22]. Compared to previous VR/AR-based remote collaboration/instruction systems, such systems included the aspect of avatar body representation. In this study, we also focused on this aspect in the context of AR-based remote instruction systems.

2.2. Role of Body Representation in Interpersonal Communication

Before exploring the effects of body representation in a remote collaboration/instruction system, we first explain the effects of body representation in interpersonal communication. Although people mainly use verbal expressions during interactions, they also employ various nonverbal behaviors to maintain interpersonal communication and collaboration [23]. Therefore, a body representation can support gestures, body orientation, proxemics, and so on. Wittaker suggested that gestures play a key role in drawing attention, turn-taking, and conducting referential tasks [17]. Heath and Luff suggested that people often respond to physical gestures automatically in face-to-face interactions [24].

Body orientation also constitutes a vital nonverbal behavior in interpersonal communication. Scheflen defined two types of body orientation—vis-à-vis orientation and parallel orientation—and indicated that people switched between these two types, depending on the situation. People used vis-à-vis orientation when the activities were related to exchange of information, such as informing or teaching, and used parallel orientation when they were engaged mutually toward a third party or object, such as discussing posters on the wall [25]. Kendon indicated that speakers change their body orientation when shifting topics in a speech [26]. Furthermore, Schegloff indicated that the upper and lower body orientations of a person, respectively, represent his/her temporal focus and dominant involvement [27].

Proxemics corresponds to the amount of space people leave between one another. Hall considered that this space is related to the relationship between the individuals involved and their manner of interaction [28].

Overall, the above-mentioned results indicate that nonverbal behaviors are important in maintaining interpersonal communication/collaboration and that body representation supports humans in effectively transferring such nonverbal behaviors.

2.3. Role of Body Representation in Remote Collaboration/Instruction

In addition to face-to-face interpersonal interactions, body representation also plays a key role in remote collaboration/instruction. Many studies have highlighted the importance of providing body information in a remote collaboration/instruction system. George et al. asked participants to receive instructions from an avatar, webcam, or voice in a memory task, and it was indicated that the social presence differed depending on the instructor design [29]. Yamamoto et al. compared an avatar-based remote instruction system with a point-based remote instruction system and indicated that, when using the avatar-based system, the task completion time was shorter and workers experienced less frustration, compared to using the pointer-based system [30]. Waldow et al. compared three types of remote instruction: face-to-face interaction, avatar-mediated interaction, and interaction with an instructor's gaze, and found that social presence had a significant difference between face-to-face interaction and interaction with an instructor's gaze and between avatar-mediated interaction and interaction with an instructor's gaze [31].

Furthermore, other researchers have also investigated the effects of body representation level of an avatar on remote collaboration. Smith et al. compared face-to-face interaction, a VR-based remote collaboration system with a hand-shaped avatar, and a VR-based remote collaboration system with a body-shaped avatar in terms of social presence. The results indicated that the body-shaped avatar increased the social presence of the interlocutor, and that participants felt they were interacting with a real person when interacting with the body-shaped avatar [1]. Yoon suggested that, in an AR-based remote collaboration system, a realistic whole-body avatar had a higher social presence than a hand-only avatar [32].

Although many studies have highlighted the importance of body representation in remote collaboration, some studies have yielded different results. The research of Chapanis et al. indicated that there was no significant difference between the task performance pertaining to face-to-face interaction and voice-only communication, which corresponds to communication without body representation [33]. The task type has been also noted to influence the effectiveness of visual information in remote collaboration. Studies have demonstrated that visual information has little impact on cognitive problem solving [34,35]. Gaver observed that, in a complex physical task, participants spent less time viewing each other [36]. Thus, an appropriate task must be selected to assess the effects of body representation.

Furthermore, in some studies, body representation was noted to have a negative impact. Doucette et al. counted the number of times local users crossed their own arms with a digital arm on a digital table. The results indicated that participants felt more awkward and tended to avoid crossing their arms when interacting with a normal-sized digital arm than a thin digital arm [2]; in other words, providing an arm body representation in a remote collaboration system may interrupt the conversation, in some cases.

As shown above, previous studies have demonstrated that body representation in a remote collaboration/instruction environment may have both positive and negative effects on the collaboration/instruction. Thus, providing a suitable body representation may optimize the quality of the remote instruction. In this regard, it is important to investigate the influence of different body representation levels on the quality of remote instruction.

Additionally, although some researchers have investigated the effects of body representation level on social presence in remote collaboration [32], insufficient research exists regarding other aspects relating to manufacturing, such as efficiency, performance, workload, usability, and preference. To this end, in this study, we compared three avatars with different body representation levels, which have been used in remote collaboration/instruction systems: avatar with whole body [1,20,32], avatar with hand and arm [21], and avatar with only hand [1,11,19,37]. Furthermore, we investigated the influence of these avatars with different body representation levels on the quality of remote instruction.

2.4. Evaluation of Remote Instruction in This Study

In this section, we describe that the factors used to evaluate the quality of remote collaboration in manufacturing: usability, workload, performance, efficiency, workload, and preference.

Usability is the degree of ease with which a system can be used to achieve the required goals. It includes several aspects, such as learnability and ease-of-use, and has been widely used to evaluate remote collaboration systems [4,5,11,19,20,38,39]. The workload is the amount of effort a user must spend to accomplish a task. No standard method to evaluate the workload in a task performed using a remote collaboration system exists; some researchers have employed a self-designed questionnaire [5], although most studies have used an existing questionnaire, such as the NASA Task Load Index (NASA-TLX) [29,30,40–42], SMEQ [20,39], or SEQ [20]. The performance is a fundamental aspect in evaluating a remote instruction system. When using a remote instruction system, one must first consider if the system can help users successfully complete the task at hand. The quality of the outcome [43] and error in the task [30] are two indicators for performance. The efficiency indicates how fast a user can accomplish a task, and its indicators include the task completion time [4,22,30,43–45] and number of tasks completed [43]. Preference indicates how much the users like a system. As the preference for a system influences the user's willingness to use the system, it has also been widely evaluated in the context of remote collaboration systems [1,20].

3. Experimental Design

3.1. Hypothesis

Previous research has indicated that body representations can support people in transferring nonverbal behavior, which is important for collaboration. Thus, in avatar-based remote instruction, we assume that an avatar with high body representation is positively associated with usability, performance, efficiency, and preference, but negatively associated with workload. Therefore, we set the following hypotheses:

Hypothesis 1 (H1). *An avatar-based remote instruction system with an avatar with a high body representation level has higher usability.*

Hypothesis 2 (H2). *In a remote instruction task, participants experience lower workload when interacting with an avatar with a high body representation level.*

Hypothesis 3 (H3). *In a remote instruction task, participants demonstrate higher performance when interacting with an avatar with a high body representation level.*

Hypothesis 4 (H4). *In a remote instruction task, participants demonstrate higher efficiency when interacting with an avatar with a high body representation level.*

Hypothesis 5 (H5). *Participants prefer to interact with an avatar with a high body representation level.*

3.2. Method

To test the five hypotheses, we conducted a within-participant design experiment with a three-level independent variable "body representation level": "Body" [1,20,32], "Hand + Arm" [21], and "Hand only" [1,11,19,37]. Further, a comparative analysis among these three conditions was performed. For each condition, we prepared an avatar that represented the corresponding body representation level (see Figure 1). The "Body" avatar was a human-like avatar, corresponding to the highest body representation level. The models of the other two avatars were the same as those of the "Body" avatar; however, the other body parts were removed. To clarify the head orientation, we attached a nose on the face of the "Body" avatar and placed a cap on its head. In addition, to avoid occluding the environment with the avatar's body, the avatar body was made semi-transparent. The movement of the avatars was prerecorded and represented differently, depending on the conditions. The details are described in Section 3.3.

3.3. Material

In this study, we chose to use prerecorded instructions (instead of real-time instructions) to minimize the differences between the conditions. We chose an order picking task [46,47], which is a common task in warehouses and manufacturing, as the task in this study. In an order picking task, a worker retrieves objects from buffer areas in response to an instructor's request. To simulate a working environment for order picking tasks, two tables (180 cm × 60 cm and 100 cm × 60 cm) were arranged in parallel (Figure 2) and 10 blocks were put on the tables. As we were interested in the relationship between the body representation level and distance between the avatar and the block, which is a factor influencing the task difficulty [48], we prepared two types of blocks to represent two levels of task difficulty: near blocks (easy) and far blocks (difficult). The near blocks were placed near the aisle, and the far blocks were placed far from the aisle. Three far blocks and three near blocks were placed on the 180 cm × 60 cm table, and two far blocks and two near blocks were placed on the 100 cm × 60 cm table.

The working environment was captured by two Intel Realsense D435 and transferred to a 3D point cloud by Unity3D. One confederate, who played the role of the instructor, was put in an empty room and wore an HTC VIVE HMD to observe the working environment remotely. In the 3D point cloud, the instructor stood in the aisle between the two tables and was asked to sequentially point to the 10 blocks five times to create five different instruction sequences. In each sequence of instructions, the instructor pointed at each block only once, where the order of pointing was defined by the experimenter. When the next block in sequence was a near block, the instructor moved to the front of the block, in order to ensure that the distance between the block and the instructor was less than arm length; this condition simulated indicating when a person could reach the block directly. When the next block was a far block, the instructor pointed at the block without moving; this condition simulated indicating when a person could not reach the block directly. We used 14 Optitrack S250e cameras to record the instruction sequences, where the duration of each sequence was approximately 30 s.

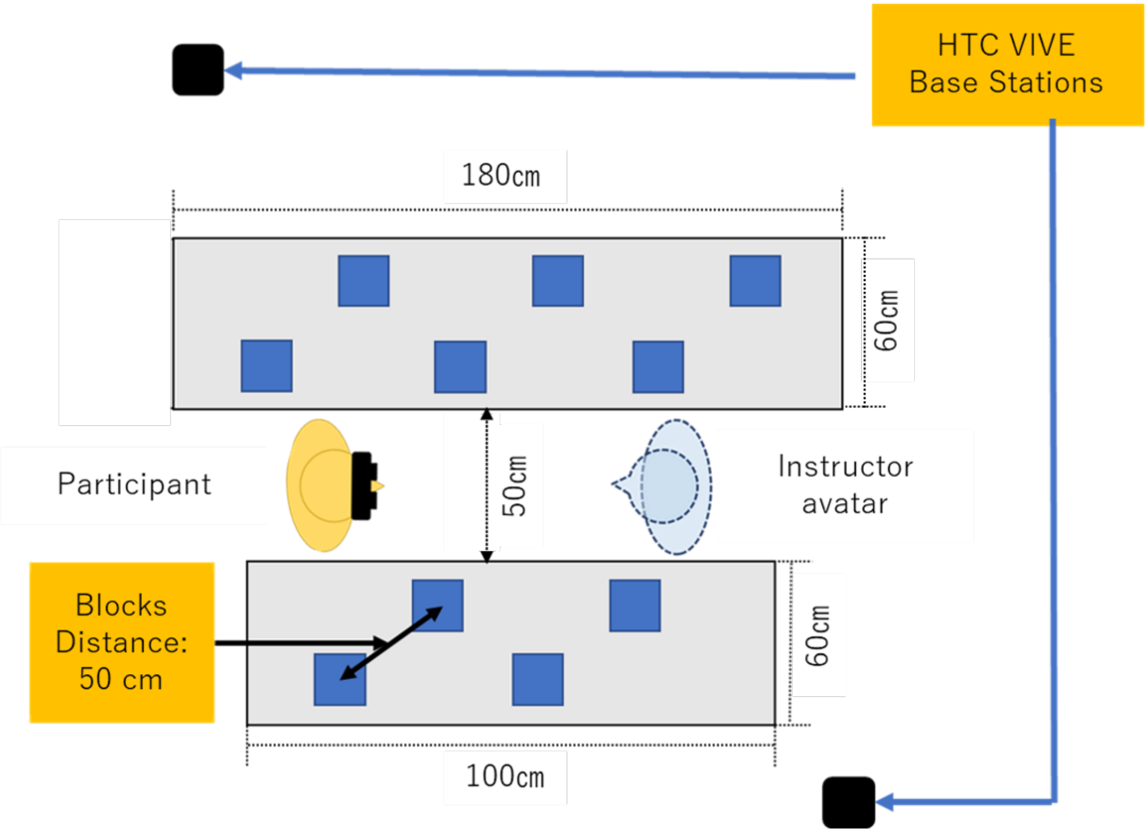

Figure 2. Experimental environment.

3.4. Participants

Ten right-handed participants (nine males and one female) from the University of Tsukuba were recruited. The average age was 23.6 and the standard deviation was 1.26. Each participant received 830 yen as compensation after the experiment. Written consent was obtained from all participants.

3.5. Procedure

The HMD of an HTC VIVE was adjusted for the best fit of each participant. The experiment consisted of three trials, and the participants stayed in the working environment (Figure 2) and interacted with different avatars in different trials. In each trial, a practice task was conducted before the main task, in which the participant was asked to stand at an initial point and close his/her eyes. At the beginning of the experiment, the participant was informed that, in the HMD, he/she would view the real world (including tables and blocks) which was captured using a stereo camera (ovrVision Pro) attached to the HMD and a virtual avatar giving him/her instructions. We asked the participant to touch the block that the avatar was pointing to as promptly as possible. When the practice task was finished, we asked the participant to close his/her eyes until the environment for the main task was prepared. In the main task, the participant repeated the same procedure, after which the experimenter asked him/her to take off the HMD and answer the questionnaires described in Section 3.6. After three trials, the participant was asked to rank the three avatars according to their preference, and a brief unstructured interview was conducted.

In terms of the experimental design, three trials involving different avatars were randomized. One sequence of instructions was chosen as the task for all the practice tasks. From the other four sequences of instructions, we randomly selected three sequences as the instructions for the three main tasks.

3.6. Measures

As mentioned above, five aspects relating to manufacturing (i.e., usability, workload, performance, efficiency, and preference) were considered for the evaluation of the avatars at the three body representation levels in an AR-based remote instruction system.

3.6.1. Usability

Although many studies have involved the use of self-designed questions to evaluate the usability of collaboration/instruction systems, we employed the system usability scale (SUS) as the questionnaire in this study. The SUS was designed by Brooke [49], and has demonstrated high reliability and validity [50]. As the SUS can be used in a wide range of systems [50,51] and previous studies have indicated that the SUS can be applied in research involving a small sample size [50,52], it was selected to evaluate the usability in this study.

The SUS is a five-point scale (0–4) with 10 items. Although the SUS has been considered to be unidimensional, Lewis and Sauro suggested that it is possible to analyze a single item to assess a specific attribute [53]. We considered the perceived complexity (Item 2), perceived ease-of-use (Item 3), perceived learnability (Item 7), and confidence-in-use (Item 9) to be related to the quality of remote instruction. Thus, in this study, we examined the effects of the body representation level on these four items.

3.6.2. Workload

To measure the workload, we selected the NASA-TLX as our evaluation technique. The NASA-TLX was developed by Hart and Staveland [54] to assess a person's subjective experience of the workload caused by a task using six sub-scales: mental demand, physical demand, temporal demand, performance, effort, and frustration. The method contains two stages: rating scales and pairwise comparison. In the rating scale stage, each sub-scale contains a statement, such as "How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?", and participants are asked to answer a rating scale consisting of 20 five-point steps, from 0 to 100. In the pairwise comparison stage, the participants compare each pair of sub-scales and judge which sub-scale was more important to the task.

3.6.3. Performance

To evaluate the performance, we measured the correctness of a task. Each task consisted of 10 sub-tasks, and the correctness was defined as the proportion of successful sub-tasks (i.e., a sub-task in which the corresponding block was correctly touched). The correctness is also associated to the understandability of the instructions and, thus, we asked the participants to respond to a self-reported question "Q1. Did you feel that it was easy to understand the instruction given by the avatar?" with a seven-point rating scale (1–7).

3.6.4. Efficiency

As described in Section 3.6, the task completion time has been widely used to assess the efficiency of systems [4,22,30,43–45]. In this work, as we used prerecorded instructions, the task completion time was fixed. Thus, in the experiment, we considered the ease of tracking an avatar and time to respond to the avatar's instruction as the two indicators associated with the efficiency.

To measure the ease of tracking an avatar, the participants were asked to respond to a self-reported question "Q2: Did you experience any difficulty in tracking the avatar during the experiment?" with a seven-point rating scale (1–7).

Regarding the time taken to respond to an avatar's instruction, we measured the time difference between the pointing gesture of the avatar and the touching gesture of the participants. As mentioned in Section 3.3, each task involved 10 sub-tasks. We constructed a model to explain the movement of both the avatar and participant in a sub-task (Figure 3). At the beginning of a sub-task, the avatar moved to the front of the block. Later, the avatar performed a pointing gesture. According to Kendon and McNeill, a gesture phase consists of three parts: preparation, holding, and retraction [55,56]. In this work, the preparation involved pointing to the block; holding involved maintaining the pointing gesture for a certain time period; and retraction involved ending the pointing gesture.

The participant first judged the correct block, based on the instruction given by the avatar. After the avatar performed holding, the participant performed a touching gesture, in which preparation involved moving his/her hand to the block; holding involved maintaining contact with the block; and retraction involved ending the touching process.

We considered the time difference between the start of the avatar's hold and the start of the participant's preparation as the time taken by participants to interpret the avatar's instruction. Note that the time difference could be negative, if the start of the participant's preparation was faster than the start of the avatar's hold.

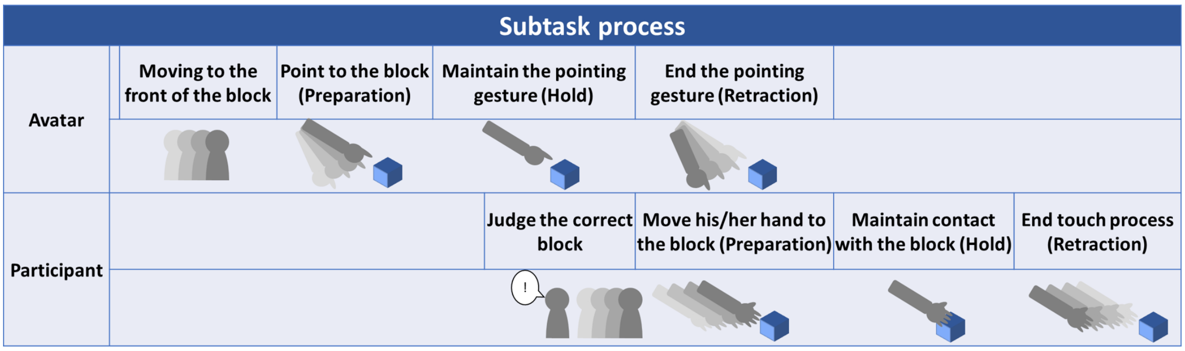
Subtask process
Avatar
Moving to the front of the block
Point to the block (Preparation)
Maintain the pointing gesture (Hold)
End the pointing gesture (Retraction)
Participant
Judge the correct block
Move his/her hand to the block (Preparation)
Maintain contact with the block (Hold)
End touch process (Retraction)
!

Figure 3. Gesture phase of a sub-task.

3.6.5. Preference

To evaluate the preference, we employed the method reported by Smith and Neff [1] and asked participants to rank the three avatars from the most to least preferred.

4. Results

4.1. Usability

Following the standard method to analyze the SUS, we first reversed the scores of even-numbered items. Next, we added the scores of 10 items and multiplied the scores by 2.5 to obtain the overall SUS score on a 100-point scale.

The overall SUS score is shown in Figure 4. A Bartlett's test suggested that the assumption of homogeneity of variances was met (Bartlett's K-squared = 0.999, $p = 0.607$) and repeated measures one-way ANOVA was used to analyze the effect of body representation level on the overall SUS score; a marginal significance was noted (F(2, 18) = 3.472, $p = 0.053$, $\eta^2_{partial} = 0.253$). Furthermore, post-hoc tests involving the Bonferroni correction were conducted. The results indicate that the overall SUS score was marginally significantly higher in the "Body" condition (M = 73.5, SD = 14.72) than in the "Hand only" condition (M = 62.25, SD = 14.55) (t = 2.56, $p = 0.059$). The "Hand + Arm" condition (M = 70.25, SD = 10.77) had no significant difference from the other two conditions.

In addition, Bartlett's tests found that the assumption of homogeneity of variances was met in Item 2: Perceived Complexity (Bartlett's K-squared = 0.484, $p = 0.785$); Item 3: Perceived Ease-of-use (Bartlett's K-squared = 2.3, $p = 0.317$); Item 7: Perceived Learnability (Bartlett's K-squared = 4.3, $p = 0.117$); and Item 9: Confidence in use (Bartlett's K-squared = 2.055, $p = 0.358$). Therefore, repeated measures one-way ANOVA was carried out to analyze the effects of body representation level on the four items. No significance between the conditions was noted for Item 2, with F(2, 18) = 0.172, $p = 0.84$, and $\eta^2_{partial} = 0.015$. A significant difference was noted for Item 3, with F(2, 18) = 3.97, $p = 0.037$, and $\eta^2_{partial} = 0.258$. The results of post-hoc tests using the Bonferroni correction indicated that the score of Item 3 was significantly higher in the "Body" condition (M = 2.9, SD = 0.88) than in the "Hand only" condition (M = 2.1, SD = 0.73) (t = 2.7, $p = 0.036$). The pairwise comparisons of the "Hand + Arm" condition (M = 2.6, SD = 0.52) with the other two conditions were not significant. For Item 7, no significance between the conditions was noted, with F(2, 18) = 1.588, $p = 0.232$, and $\eta^2_{partial} = 0.138$. For Item 9 (confidence-in-use), a significant difference was observed, with F(2, 18) = 7.13, $p = 0.005$, and $\eta^2_{partial} = 0.423$. The results of post-hoc tests using the Bonferroni correction indicated that the score of Item 9 was significantly higher in the "Body" condition (M = 3.1, SD = 0.74) than in the "Hand only" condition (M = 2.1, SD = 1.1) (t = 3.569, $p = 0.007$). The score in the "Hand + Arm" condition (M = 2.9, SD = 1.2) was significantly higher than that in the "Hand only" condition (t = 2.855, $p = 0.032$). The pairwise comparison of the "Body" condition with the "Hand + Arm" condition was not significant.

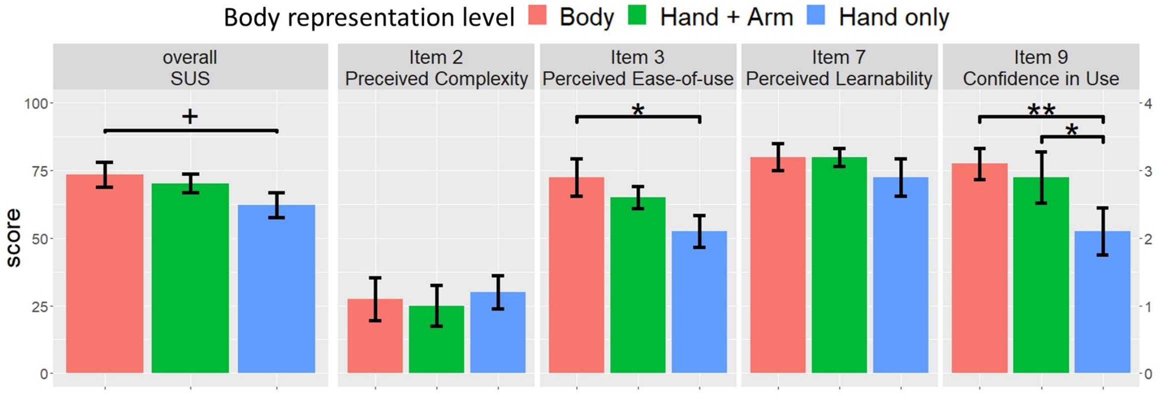

Figure 4. Overall SUS score and individual item scores (+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$). All error bars represent standard errors.

4.2. Workload

To analyze the workload using the NASA-TLX, we first analyzed each sub-scale of the workload. Each sub-scale met the assumption of homogeneity of variances (Table 1) and repeated measures one-way ANOVA was used to analyze the effect of the body representation level on each sub-scale of the workload (Figure 5). For the physical demand, a marginally significant difference was observed, with F(2, 18) = 2.919, p = 0.08, and $\eta^2_{partial} = 0.235$, and the results of the post-hoc test involving the Bonferroni correction indicated that the score of the physical demand in the "Body" condition (M = 48, SD = 24.74) was marginally significantly lower than that in the "Hand only" condition (M = 61, SD = 25.36) (t = −2.307, p = 0.099). The score of the physical demand in the "Hand + Arm" condition (M = 51, SD = 23.07) did not exhibit a significant difference with the scores in the other two conditions. For the temporal demand, a significant effect was noted, with F(2, 18) = 7.205, p = 0.005, and $\eta^2_{partial} = 0.433$. Results of the post-hoc test involving the Bonferroni correction indicate that the score of the temporal demand in the "Body" condition (M = 50.5, SD = 26.92) was significantly lower than that in the "Hand only" condition (M = 68.5, SD = 20.69) (t = −3.543, p = 0.007). The score of the temporal demand in the "Hand + Arm" condition (M = 53.5, SD = 22.12) was significantly lower than that in the "Hand only" condition (t = −2.952, p = 0.026). Pairwise comparison between the "Body" and "Hand + Arm" conditions did not indicate any significant effect on the score of the temporal demand. In terms of the other four sub-scales, the three conditions did not have a significant effect on the scores of mental demand (F(2, 18) = 0.853, p = 0.443, $\eta^2_{partial} = 0.083$), performance (F(2, 18) = 1.283, p = 0.301, $\eta^2_{partial} = 0.117$), effort (F(2, 18) = 0.158, p = 0.855, $\eta^2_{partial} = 0.017$), and frustration (F(2, 18) = 1.418, p = 0.268, $\eta^2_{partial} = 0.122$).

Furthermore, we summed up the six subscales of the NASA-TLX with their weights to obtain the overall NASA-TLX score. The overall NASA-TLX score met the assumption of homogeneity of variances (Table 1) and we used a one-way ANOVA to analyze the effect of the body representation level on the overall NASA-TLX score (Figure 5, Overall). The results indicate a significant effect on the overall NASA-TLX score (F(2, 18) = 5.702, p = 0.012, $\eta^2_{partial} = 0.384$). Results of the post-hoc test with the Bonferroni correction indicate that the overall NASA-TLX score in the "Body" condition (M = 49.6, SD = 17.946) was significantly lower than that in the "Hand only" condition (M = 58.78, SD = 20.858) (t = −3.346, p = 0.011). The "Hand + Arm" condition (M = 53.106, SD = 19.443) did not have a significant difference with the other two conditions.

Table 1. Result of Bartlett's test for NASA-TLX.

	Bartlett's K-Squared	***p* Value**
Mental Demand	0.79	0.673
Physical Demand	0.082	0.96
Temporal Demand	0.663	0.72
Performance	0.371	0.83
Effort	0.506	0.78
Frustration	0.022	0.99
Overall	0.193	0.91

4.3. Performance

We used two-way repeated measures ANOVA to test the effects of the body representation level and block distance (i.e., far or near blocks) (Figure 6, left) on the correctness. The individual difference of each participant was considered as a random factor and introduced into the model. A significant effect of block distance on the correctness was noted (F(1, 45) = 26.328, $p < 0.001$); however, the body representation level did not exert a significant effect on the correctness (F(2, 45) = 1.041, p = 0.361). In addition, no significant interaction between the body representation level and block distance was observed (F(2, 45) = 0.235, p = 0.792).

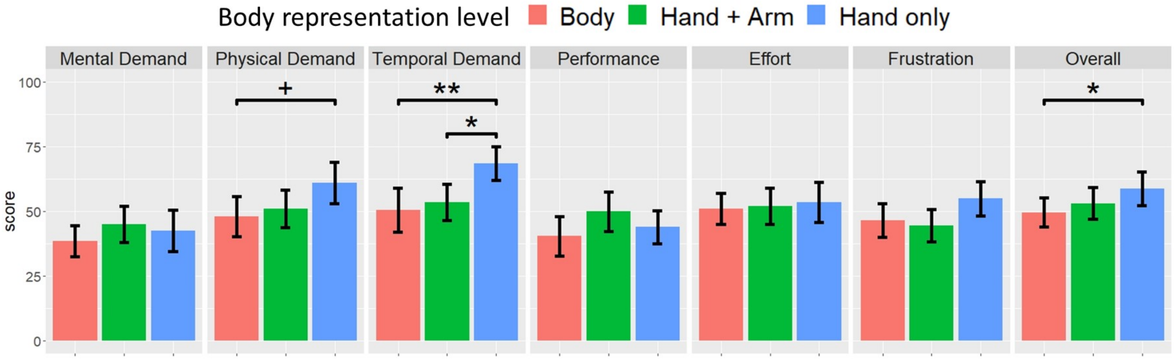

Figure 5. Score of NASA-TLX sub-scales (+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$). All error bars represent standard errors.

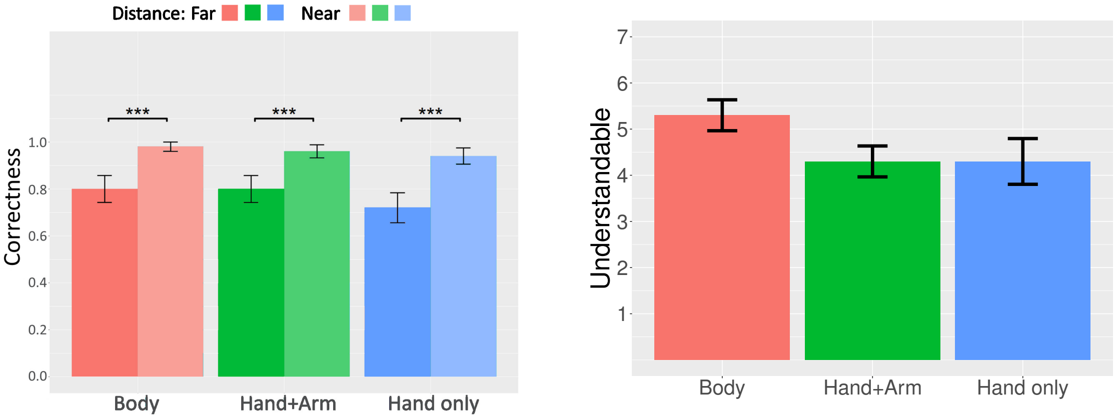

Figure 6. Left: result of correctness; right: result of Q1 (Understandability). All error bars represent standard errors (*** $p < 0.001$).

In addition, we used repeated measures one-way ANOVA to analyze the effect of body representation level on the score of Q1 (Figure 6, right). The result of Bartlett's test found that the assumption of homogeneity of variances was met (Bartlett's K-squared = 1.869, $p = 0.39$). The results indicate that there existed a marginally significant effect on the score of Q1 (F(2, 18) = 3.103, $p = 0.07$, $\eta^2_{partial} = 0.222$). However, post-hoc tests with Bonferroni correction suggested that no significant differences existed between each pair of conditions among the "Body" (M = 5.3, SD = 1.06), "Hand + Arm" (M = 4.3, SD = 1.06), and "Hand only" (M = 4.3, SD = 1.57) conditions.

4.4. Efficiency

As Q2 was a negative item, we first reversed its score. The score met the assumption of homogeneity of variances (Bartlett's K-squared = 0.022, $p = 0.99$). Then, we used repeated measures one-way ANOVA to test the effect of body representation level on the score of Q2 (Figure 7, left). The results indicate that the body representation level exerted a significant influence on the score of Q2 (F(2, 18) = 5.472, $p = 0.014$, $\eta^2_{partial} = 0.289$). The results of post-hoc tests involving the Bonferroni correction indicate that the score of Q2 in the "Body" condition (M = 5.1, SD = 1.52) was significantly higher than that in the "Hand only" condition (M = 2.9, SD = 1.45) (t = 3.308, $p = 0.012$). However, the score of Q2 in the "Hand + Arm" condition (M = 4, SD = 1.49) did not exhibit a significant difference from the scores of Q2 in the other two conditions.

Before analyzing the time difference between the start of the avatar's hold and that of the participant's preparation, we first asked two raters to annotate the preparation of the participant's touching gestures (Figure 3) in 90 data items. Each hand movement consisted of acceleration and deceleration. We defined the last hand movement before the participants touched the block as the preparation part of the participant's touch gestures. Later, we calculated the inter-rater reliability using the intra-class correlation coefficient (ICC). The ICC form was as follows: "two-way random effects, absolute agreement, and single rater" [57,58]. Two data items were removed because the participant touched the wrong block. A high degree of reliability was found between the two raters. The average measured ICC was 0.938 with a 95% confidence interval from 0.898–0.961 (F(87,48) = 34.8, $p < 0.001$). Subsequently, we asked one rater to annotate the other 210 data items.

The result of Levene's median test show that the time difference met the assumption of homogeneity of variance (F(59, 231) = 0.968, $p = 0.547$) and we built a linear mixed model to predict the time difference based on the block distance and body representation level (Figure 7, right). The individual differences of participants were considered as a random factor and introduced into the model. Nine data points were removed because the participant either gave up on touching the block, touched the wrong block, or touched more than one block.

The results indicate that both the block distance (F(1, 276) = 62.473, $p < 0.001$, $\eta^2_{partial} = 0.183$) and body representation level (F(2, 276) = 4.392, $p = 0.013$, $\eta^2_{partial} = 0.031$) exerted significant influences on the time difference. In addition, no significant interaction between the block distance and body representation level was noted (F(2, 276) = 1.462, $p = 0.234$, $\eta^2_{partial} = 0.001$). Furthermore, a post-hoc test involving the Bonferroni correction was applied to test the effect of different body representation levels on the time difference. The results indicate that the time difference was significantly lower in the "Body" condition (M = 0.15, SD = 0.333) than in the "Hand only" condition (M = 0.279, SD = 0.392) (t = -2.761, $p = 0.018$). The time difference in the "Hand + Arm" condition (M = 0.18, SD = 0.335) was also marginally significantly lower than in the "Hand only" condition (t = -2.333, $p = 0.061$). However, no significant difference between the "Hand + Arm" and "Hand only" conditions was noted.

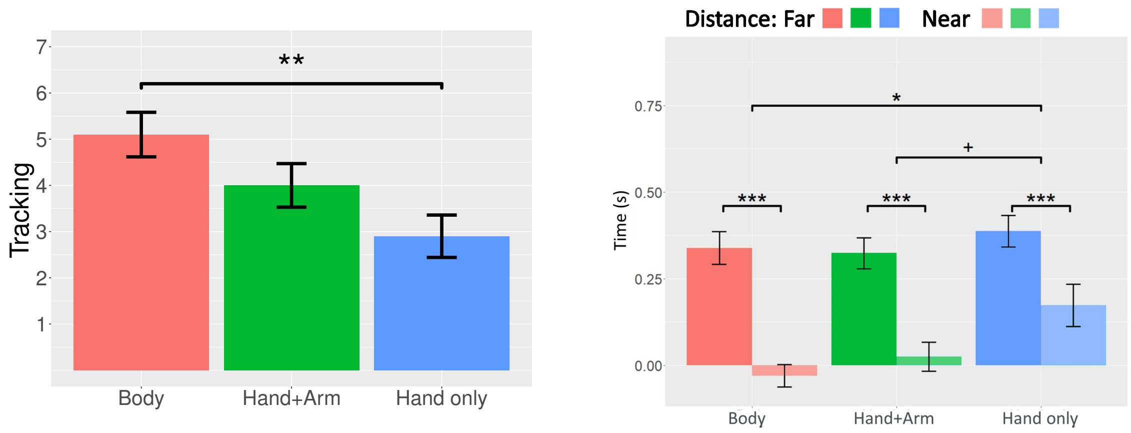

Figure 7. (**Left**) Result of Q2 (Tracking); and (**right**) result of time difference. All error bars represent standard errors.

4.5. Preference

A non-parametric Friedman test was conducted to analyze the ranked data of preference of avatars with three body representation levels. No significant difference between the avatars was noted ($\chi^2(2) = 2.4$, $p = 0.301$) (Figure 8).

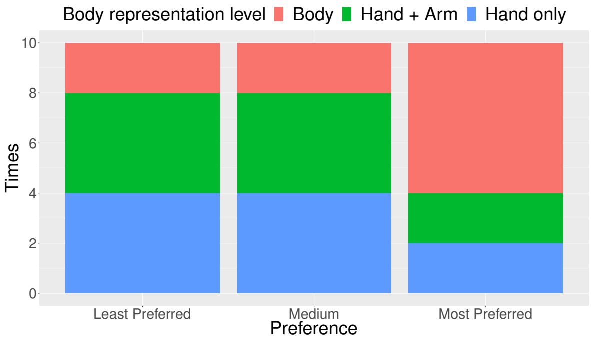

Figure 8. Preference of the three types of avatars.

5. Discussion

5.1. Usability of Avatars with Different Body Representation Levels

Hypothesis H1 (an avatar-based remote instruction system with an avatar with a high body representation level has higher usability) was supported, as the overall SUS scores exhibited a marginally significant difference among the three body representation levels and the overall SUS scores in the "Body" condition were marginally higher than in the "Hand only" condition. In addition to relative comparison, the overall SUS score could be interpreted using Sauro's grading scale [50], according to which "Body" had a "B-" grade (65–69%), "Hand + Arm" had a "C" grade (41–59%), and "Hand only" had a "D" grade (15–34%). The "B-" and "C" grades indicate that the usabilities of "Body" and "Hand + Arm" were acceptable. However, the "D" grade indicates that the overall SUS score of "Hand only" was lower than 68 (the mean of the overall SUS score), thereby indicating that the usability of the "Hand only" avatar was less than acceptable [50,59].

We further considered the sub-dimensions of usability. The perceived ease-of-use and confidence-in-use demonstrated a significant difference between the "Body" and "Hand only" conditions. In contrast, the perceived complexity and perceived learnability exhibited no significant differences among the three conditions. We further compared the mean scores of these two sub-dimensions, based on the item benchmarks defined by Lewis et al. [53]. The mean values of the perceived complexity of the avatars with the three body representation levels were less than 1.44 (the scores of SUS items were 1–5 in the research of Lewis et al. and the scores of SUS items in this research were 0–4; thus, we shifted their benchmark), which was the mean score of Item 2. The mean values of the perceived learnability were all higher than 2.71 (the benchmark was shifted), which was the mean score of Item 7. Based on these results, it can be considered that the participants found it easy to learn how to use all the three types of avatars, as they were not complex. Considering these aspects, we believe H1 to be partially supported.

5.2. Quality of Instruction with Different Avatars

The results of NASA-TLX supported H2 (in a remote instruction task, participants experience lower workload when interacting with an avatar with a high body representation level). Among the six sub-scales of NASA-TLX, no significant difference among the body representation levels was noted, except in the temporal and physical demand sub-scales. To explain this phenomenon, we considered that both sub-scales were associated with the difficulty of avatar tracking. The results of Q2 indicated that the "Body" avatar was easier for participants to track than the "Hand only" avatar. In addition, we observed that participants often failed to track the "Hand only" avatar, and tended to twist their bodies to search for it. The effort and time required in searching for the "Hand only" avatar caused this difference among the body representation levels, in terms of the temporal and physical demand sub-scales.

Regarding H3 (in a remote instruction task, participants demonstrate higher performance when interacting with an avatar with a high body representation level), based on the results of Q1 and correctness (proportion of successfully performed subtasks), it was noted that all the three avatars could correctly transfer the instructions. The only difference pertaining to performance was in terms of the block distance (far or near blocks) influencing the accuracy of the task. The finding that H3 was not supported in this study was unexpected. Previous research has indicated that the pointing person put a finger between the target and eye to create a line; however, the observer estimated the pointing target based on the extrapolation of the arm–finger line, which caused misinterpretation [60]. Another VR collaboration research proposed warping a virtual character's arm to match the observer's perception of pointing gestures, which could reduce the misinterpretation of pointing [61]. The following are some possible reasons the arm did not play an important role in this study: Previous research has demonstrated that distance was a significant factor leading to misinterpretation of pointing [48]. However, in most situations considered in this study, the distance between the avatar and block was less than 1.5 m. Under similar conditions, Bangerter et al. indicated that the mean horizontal error of pointing interpretation was approximately 12 cm [62], which was considerably smaller than the distances between the blocks (50 cm) used in this study. Thus, the possibility of misunderstanding might be low in all conditions. Furthermore, the location of participants is important to ensure the accuracy of pointing interpretation. In the collaborative VR study of Wong, it has been reported that the accuracy of pointing is higher when the observer was standing behind the pointer, compared to when the observer is standing beside the pointer [63]. In this study, due to the environment, participants had several opportunities to stand behind the avatar and observe the instructions, resulting in a higher accuracy of the pointing interpretation. Considering the combined effect of these aspects, the performance did not exhibit a significant difference among the body representation levels in this study.

H4 (in a remote instruction task, participants demonstrate higher efficiency when interacting with an avatar with a high body representation level) was considered to be supported, based on the results of both the ease of tracking the avatar and time to respond to the avatar's instruction. In terms of the ease of tracking, the results of Q2 indicated that participants felt that tracking a "Body" avatar was significantly easier than tracking a "Hand only" avatar, likely due to the large size of the avatar.

In terms of the time taken to response to an avatar's instruction, the results of the time difference between the start of the avatar's hold and start of participants' preparation indicated that participants spent less time in responding to the instructions regarding near blocks than those of the far blocks and to the instructions of the "Body" avatar, compared to those of the "Hand only" avatar. Regarding the former result, as assumed, the distance between the hand and block influenced the time for participants to respond to the instruction; for the participants, the instructions pertaining to the near blocks were clearer than those for the far blocks. Regarding the latter result, a likely reason participants spent less time in responding to the "Body" avatar's instruction is that the "Body" avatar consisted of body orientation. As a person's body orientation is related to his/her current involvement [27], participants likely automatically judged the instructions based on the avatar's body orientation, resulting in a faster response time.

Another interesting finding is that the time difference between the start of the avatar's hold and the start of participants' preparation was negative in the "Body" condition, which meant the participants judged the instructions even before the instructions were given. This provides more evidence that the body orientation facilitated participants in predicting and judging the instructions.

5.3. Preference of Different Avatars

Regarding H5 (participants prefer to interact with an avatar with a high body representation level), no significant difference among the preference of the three body representation levels was noted and, thus, H5 was not considered to be supported. However, several interesting comments were noted in the interview.

Six participants appeared to prefer the "Body" avatar most, and felt that the "Hand only" and "Hand + Arm" avatars were peculiar. Participants 3 and 9 mentioned that it was unnatural to see two arms floating in space. In contrast, the participants appeared to like the resemblance to reality of the "Body" avatar. Participant 9 said "I think arm avatar was sufficient for this experiment; however, I preferred to interact with the body avatar because it felt like I was interacting with a real human." These results were similar to those reported by Yoon, in which participants preferred the body avatar more than the hand avatar or upper body avatar and disliked the hand avatar the most, owing to it not having a body [32]. Participant 2 further mentioned that the motion of the "Body" avatar, including bobbing (body moving up and down during each step) and lunging (moments of forwarding acceleration, sometimes associated with the forward thrust of taking each step), made the avatar more realistic.

Surprisingly, some participants who ranked the "Body" low in preference also mentioned that they felt like they were interacting with a real human when interacting with the "Body" avatar. However, they reported that it was not pleasant when the "Body" avatar walked through their bodies. From the video, we could observe that these participants tended to move backward when the avatar was walking toward them. This opinion and behavior may be attributed to the need for maintenance of personal space by humans [28]. Other studies also demonstrated that people maintained their personal space when interacting with an avatar in a virtual environment [64]. Additionally, in this study, the avatar did not only walk around the participants, but also walked through the participants. Andersen et al. indicated that touch avoidance is a natural behavior for humans [65]. Doucette also noted that people avoided touching a digital arm during remote collaboration [2].

5.4. Limitations and Future Work

This study demonstrated that an avatar with a high body representation level had a higher usability and that participants experienced a lower workload and higher efficiency when interacting with it. However, we chose to use a prerecorded instruction instead of real-time instruction. Some studies have explored the differences between prerecorded VR and real-time VR. Borst et al. suggested that, in a teacher-guided educational VR system, the understandability of a teacher's pointing was significantly different between live guidance and prerecorded guidance [38]. Therefore, the quality of remote instruction may differ between prerecorded instruction and real-time instruction. Further research investigating the effect of body representation level in a real-time AR-based remote instruction system is, thus, necessary.

Additionally, in this study, we selected an order picking task as the main task for the experiment. In an order picking task, participants tend to spend more time viewing the avatar, which may enhance the effect of body representation. However, there are many different tasks in manufacturing. The effect of body representation level might differ when performing other tasks. As further work, the study must be repeated with different types of task, such as assembly tasks.

Regarding the experimental design, we had a relatively small sample size and all participants were students. Workers with different expertise levels might have different reactions to avatars in remote instruction. Thus, to clarify the effects of body representation in AR-based remote instruction, further research in practical situations is necessary.

Furthermore, to completely represent the full avatar, we used a video see-through AR device, ovrVision Pro, which had a wider view angle than the modern optical see-through AR device (i.e., Hololens) in this study. However, the view angle of ovrVision Pro is still smaller than that of the human eye. Thus, the effect of view angle on quality of AR-based remote instruction should be investigated in the future. Additionally, in our experiment, to avoid the impact of the low resolution caused by the ovrVision Pro, we chose appropriate blocks which were big enough to be seen through the ovrVision Pro, and no participant mentioned the resolution issue in the interview. However, there are various sizes of items in manufacturing; thus, a high-resolution video see-through AR device or optical see-through AR device should be used in future work.

As mentioned in Section 5.3, participants felt uncomfortable when the full body avatar walked through them. As a future work, it is important to consider the invasion of personal space in an avatar-based remote instruction system to achieve high quality remote instruction.

6. Conclusions

In this study, we investigated the effects of body representation level in an AR-based remote instruction system by designing three types of avatars (whole body, hand and arm, and hand only) and performed a comparative analysis. The results indicate that the usability of an AR-based remote instruction system having an avatar with a whole body was higher than that with a hand-only avatar. In addition, the participants experienced a lower workload and higher efficiency when interacting with an avatar with a whole body.

Author Contributions: Conceptualization, T.-Y.W., M.O., H.K., and Y.S.; Data curation, T.-Y.W., and Y.S.; Formal analysis, T.-Y.W., Y.S., and M.O.; Funding acquisition, M.O., H.K., and Y.S.; Investigation, T.-Y.W.; Methodology, T.-Y.W.; Project administration, H.K.; Software, T.-Y.W. and Y.S.; Supervision, T.-Y.W., M.O., and H.K.; Validation, M.O. and H.K.; Visualization, T.-Y.W.; Writing—original draft, T.-Y.W.; and Writing—review and editing, T.-Y.W., M.O., H.K., and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI (Grant Number JP16K16097).

Acknowledgments: The authorswould like to express their very great appreciation to Chi-Lan Yang and Ikkaku Kawaguchi for giving valuable writing advice.

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

References

1. Smith, H.J.; Neff, M. Communication Behavior in Embodied Virtual Reality. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 289:1–289:12. [CrossRef]
2. Doucette, A.; Gutwin, C.; Mandryk, R.L.; Nacenta, M.; Sharma, S. Sometimes when We Touch: How Arm Embodiments Change Reaching and Collaboration on Digital Tables. In Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW '13, San Antonio, TX, USA, 23–27 February 2013; ACM: New York, NY, USA, 2013; pp. 193–202. [CrossRef]
3. Shu, L.; Flowers, W. Groupware Experiences in Three-dimensional Computer-aided Design. In Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work, CSCW '92, Toronto, ON, Canada, 31 October–4 November 1992; ACM: New York, NY, USA, 1992; pp. 179–186. [CrossRef]
4. Tait, M.; Billinghurst, M. The Effect of View Independence in a Collaborative AR System. *Comput. Support. Coop. Work (CSCW)* **2015**, *24*, 563–589. [CrossRef]
5. Piumsomboon, T.; Lee, Y.; Lee, G.; Billinghurst, M. CoVAR: A Collaborative Virtual and Augmented Reality System for Remote Collaboration. In Proceedings of the SIGGRAPH Asia 2017 Emerging Technologies, SA '17, Bangkok, Thailand, 27–30 November 2017; ACM: New York, NY, USA, 2017; pp. 3:1–3:2. [CrossRef]
6. Piumsomboon, T.; Day, A.; Ens, B.; Lee, Y.; Lee, G.; Billinghurst, M. Exploring enhancements for remote mixed reality collaboration. In Proceedings of the SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications, SA '17, Bangkok, Thailand, 27–30 November 2017; ACM: New York, NY, USA, 2017; p. 16.
7. Pausch, R.; Pausch, R.; Proffitt, D.; Williams, G. Quantifying Immersion in Virtual Reality. In Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '97, Los Angeles, CA, USA, 3–8 August 1997; ACM Press; Addison-Wesley Publishing Co.: New York, NY, USA, 1997; pp. 13–18. [CrossRef]
8. Ruddle, R.A.; Payne, S.J.; Jones, D.M. Navigating Large-Scale Virtual Environments: What Differences Occur Between Helmet-Mounted and Desk-Top Displays? *Presence* **1999**, *8*, 157–168. [CrossRef]
9. Ragan, E.D.; Kopper, R.; Schuchardt, P.; Bowman, D.A. Studying the Effects of Stereo, Head Tracking, and Field of Regard on a Small-Scale Spatial Judgment Task. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 886–896. [CrossRef]

10. Schuchardt, P.; Bowman, D.A. The Benefits of Immersion for Spatial Understanding of Complex Underground Cave Systems. In Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology, VRST '07, Newport Beach, CA, USA, 5–7 November 2007; ACM: New York, NY, USA, 2007; pp. 121–124. [CrossRef]
11. Huang, W.; Alem, L.; Tecchia, F.; Duh, H.B.L. Augmented 3D hands: A gesture-based mixed reality system for distributed collaboration. *J. Multimodal User Interfaces* **2018**, *12*, 77–89. [CrossRef]
12. Gergle, D.; Kraut, R.E.; Fussell, S.R. Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. *J. Lang. Soc. Psychol.* **2004**, *23*, 491–517. [CrossRef]
13. Gergle, D.; Kraut, R.E.; Fussell, S.R. Using Visual Information for Grounding and Awareness in Collaborative Tasks. *Hum. Comput. Interact.* **2013**, *28*, 1–39. [CrossRef]
14. Fussell, S.R.; Setlock, L.D.; Yang, J.; Ou, J.; Mauer, E.; Kramer, A.D.I. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Hum. Comput. Interact.* **2004**, *19*, 273–309._3. [CrossRef]
15. Gergle, D.; Rose, C.P.; Kraut, R.E. Modeling the Impact of Shared Visual Information on Collaborative Reference. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07, San Jose, CA, USA, 28 April–3 May 2007; ACM: New York, NY, USA, 2007; pp. 1543–1552. [CrossRef]
16. Kraut, R.E.; Gergle, D.; Fussell, S.R. The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-presence. In Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, CSCW '02, New Orleans, LA, USA, 16–20 November 2002; ACM: New York, NY, USA, 2002; pp. 31–40. [CrossRef]
17. Whittaker, S. Theories and methods in mediated communication. In *Handbook of Discourse Processes*; Graesser, A., Gernsbacher, M., Goldman, S., Eds.; Erlbaum: Mahwah, NJ, USA, 2003; pp. 253–293.
18. Higuchi, K.; Yonetani, R.; Sato, Y. Can Eye Help You? Effects of Visualizing Eye Fixations on Remote Collaboration Scenarios for Physical Tasks. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, San Jose, CA, USA, 7–12 May 2016; ACM: New York, NY, USA, 2016; pp. 5180–5190.
19. Huang, W.; Alem, L.; Tecchia, F. HandsIn3D: Supporting Remote Guidance with Immersive Virtual Environments. In Proceedings of the Human-Computer Interaction—INTERACT 2013, Cape Town, South Africa, 2–6 September 2013; Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 70–77.
20. Piumsomboon, T.; Lee, G.A.; Hart, J.D.; Ens, B.; Lindeman, R.W.; Thomas, B.H.; Billinghurst, M. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, Montreal, QC, Canada, 21–26 April 2018; ACM: New York, NY, USA, 2018; pp. 46:1–46:13. [CrossRef]
21. Kolkmeier, J.; Harmsen, E.; Giesselink, S.; Reidsma, D.; Theune, M.; Heylen, D. With a Little Help from a Holographic Friend: The OpenIMPRESS Mixed Reality Telepresence Toolkit for Remote Collaboration Systems. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST '18, Tokyo, Japan, 28 November–1 December 2018; ACM: New York, NY, USA, 2018; pp. 26:1–26:11. [CrossRef]
22. Pejsa, 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 & Social Computing, CSCW '16, San Francisco, CA, USA, 27 February–2 March 2016; ACM: New York, NY, USA, 2016; pp. 1716–1725. [CrossRef]
23. Mehrabian, A. *Nonverbal Communication*; Aldine Publishing Company: Chicago, IL, USA, 1972.
24. Heath, C.; Luff, P. Disembodied Conduct: Communication Through Video in a Multi-media Office Environment. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '91, New Orleans, LA, USA, 27 April–2 May 1991 ; ACM: New York, NY, USA, 1991; pp. 99–103. [CrossRef]
25. Scheflen, A.E. The Significance of Posture in Communication Systems. *Psychiatry* **1964**, *27*, 316–331, doi:10.1080/00332747.1964.11023403. [CrossRef]
26. Kendon, A. Chapter 9—Some Relationships Between Body Motion and Speech: An Analysis of an Example. In *Studies in Dyadic Communication*; Pergamon General Psychology Series; Siegman, A.W., Pope, B., Eds.; Pergamon: Amsterdam, The Netherlands, 1972; Volume 7, pp. 177–210. [CrossRef]
27. Schegloff, E.A. Body Torque. *Soc. Res.* **1998**, *65*, 535–596.
28. Hall, E. *The Hidden Dimension*; Anchor Books: Garden City, NY, USA, 1992.

29. George, C.; Spitzer, M.; Hussmann, H. Training in IVR: Investigating the Effect of Instructor Design on Social Presence and Performance of the VR User. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST '18, Tokyo, Japan, 28 November–1 December 2018; ACM: New York, NY, USA, 2018; pp. 27:1–27:5. [CrossRef]
30. Yamamoto, T.; Otsuki, M.; Kuzuoka, H.; Suzuki, Y. Tele-Guidance System to Support Anticipation during Communication. *Multimodal Technol. Interact.* **2018**, *2*, 55. [CrossRef]
31. Waldow, K.; Fuhrmann, A.; Grünvogel, S.M. Investigating the Effect of Embodied Visualization in Remote Collaborative Augmented Reality. In *Virtual Reality and Augmented Reality*; Bourdot, P., Interrante, V., Nedel, L., Magnenat-Thalmann, N., Zachmann, G., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 246–262.
32. Yoon, B.; Kim, H.I.; Lee, G.; Billinghurst, M.; Woo, W. The Effect of Avatar Appearance on Social Presence in an Augmented Reality Remote Collaboration. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019.
33. Chapanis, A.; Ochsman, R.B.; Parrish, R.N.; Weeks, G.D. Studies in Interactive Communication: I. The Effects of Four Communication Modes on the Behavior of Teams During Cooperative Problem-Solving. *Hum. Factors* **1972**, *14*, 487–509. [CrossRef]
34. Williams, E. Experimental comparisons of face-to-face and mediated communication: A review. *Psychol. Bull.* **1977**, *84*, 963. [CrossRef]
35. O'Malley, C.; Langton, S.; Anderson, A.; Doherty-Sneddon, G.; Bruce, V. Comparison of face-to-face and video-mediated interaction. *Interact. Comput.* **1996**, *8*, 177–192. [CrossRef]
36. Gaver, W.W.; Sellen, A.; Heath, C.; Luff, P. One is Not Enough: Multiple Views in a Media Space. In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, CHI '93, Amsterdam, The Netherlands, 24–29 April 1993; ACM: New York, NY, USA, 1993; pp. 335–341. [CrossRef]
37. Sodhi, R.S.; Jones, B.R.; Forsyth, D.; Bailey, B.P.; Maciocci, G. BeThere: 3D Mobile Collaboration with Spatial Input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, Paris, France, 27 April–2 May 2013; ACM: New York, NY, USA, 2013; pp. 179–188. [CrossRef]
38. Borst, C.W.; Lipari, N.G.; Woodworth, J.W. Teacher-Guided Educational VR: Assessment of Live and Prerecorded Teachers Guiding Virtual Field Trips. In Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, Germany, 18–22 March 2018; pp. 467–474. [CrossRef]
39. Kim, S.; Lee, G.; Huang, W.; Kim, H.; Woo, W.; Billinghurst, M. Evaluating the Combination of Visual Communication Cues for HMD-based Mixed Reality Remote Collaboration. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, Glasgow, Scotland, UK, 4–9 May 2019; ACM: New York, NY, USA, 2019; pp. 173:1–173:13. [CrossRef]
40. Pouliquen-Lardy, L.; Milleville-Pennel, I.; Guillaume, F.; Mars, F. Remote collaboration in virtual reality: Asymmetrical effects of task distribution on spatial processing and mental workload. *Virtual Real.* **2016**, *20*, 213–220. [CrossRef]
41. Tan, C.S.S.; Schöning, J.; Luyten, K.; Coninx, K. Investigating the Effects of Using Biofeedback As Visual Stress Indicator During Video-mediated Collaboration. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, Toronto, ON, Canada, 26 April–1 May 2014; ACM: New York, NY, USA, 2014; pp. 71–80. [CrossRef]
42. Aschenbrenner, D.; Leutert, F.; Çençen, A.; Verlinden, J.; Schilling, K.; Latoschik, M.; Lukosch, S. Comparing Human Factors for Augmented Reality Supported Single-User and Collaborative Repair Operations of Industrial Robots. *Front. Robot. AI* **2019**, *6*, 37. [CrossRef]
43. Kraut, R.E.; Miller, M.D.; Siegel, J. Collaboration in Performance of Physical Tasks: Effects on Outcomes and Communication. In Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, CSCW '96, Boston, MA, USA, 16–20 November 1996; ACM: New York, NY, USA, 1996; pp. 57–66. [CrossRef]
44. Kiyokawa, K.; Takemura, H.; Yokoya, N. A collaboration support technique by integrating a shared virtual reality and a shared augmented reality. In Proceedings of the IEEE SMC '99 International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028), Tokyo, Japan, 12–15 October 1999; Volume 6, pp. 48–53. [CrossRef]

45. Kuzuoka, H. Spatial Workspace Collaboration: A SharedView Video Support System for Remote Collaboration Capability. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '92, Monterey, CA, USA, 3–7 May 1992; ACM: New York, NY, USA, 1992; pp. 533–540. [CrossRef]
46. Weaver, K.A.; Baumann, H.; Starner, T.; Iben, H.; Lawo, M. An Empirical Task Analysis of Warehouse Order Picking Using Head-mounted Displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, Atlanta, GA, USA, 10–15 April 2010; ACM: New York, NY, USA, 2010; pp. 1695–1704. [CrossRef]
47. Funk, M.; Mayer, S.; Nistor, M.; Schmidt, A. Mobile in-situ pick-by-vision: Order picking support using a projector helmet. In Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Corfu Island, Greece, 29 June–1 July 2016; ACM: New York, NY, USA, 2016; p. 45.
48. Herbort, O.; Kunde, W. How to point and to interpret pointing gestures? Instructions can reduce pointer–observer misunderstandings. *Psychol. Res.* **2018**, *82*, 395–406. [CrossRef]
49. Brooke, J. SUS: A quick and dirty usability scale. In *Usability Evaluation in Industry*; Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B., Eds.; Taylor & Francis: Bristol, PA, USA, 11 June 1996; pp. 189–194.
50. Sauro, J.; Lewis, J.R. *Quantifying the User Experience: Practical Statistics for User Research*, 1st ed.; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2012.
51. Brooke, J. SUS: A Retrospective. *J. Usability Stud.* **2013**, *8*, 29–40.
52. Tullis, T.S.; Stetson, J.N. A comparison of questionnaires for assessing website usability. In Proceedings of the Usability Professionals Association (UPA) 2004 Conference, Minneapolis, MN, USA, 7–11 June 2004.
53. Lewis, J.R.; Sauro, J. Item Benchmarks for the System Usability Scale. *J. Usability Stud.* **2018**, *13*, 158–167.
54. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*; Elsevier: Amsterdam, The Netherlands, 1988; Volume 52, pp. 139–183.
55. Kendon, A. *Gesture: Visible Action as Utterance*; Cambridge University Press: Cambridge, UK, 2004.
56. McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*; Hand and Mind: What Gestures Reveal about Thought; University of Chicago Press: Chicago, IL, USA, 1992.
57. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef]
58. McGraw, K.O.; Wong, S.P. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* **1996**, *1*, 30. [CrossRef]
59. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum. Comput. Interact.* **2008**, *24*, 574–594, doi:10.1080/10447310802205776. [CrossRef]
60. Herbort, O.; Kunde, W. Spatial (mis-) interpretation of pointing gestures to distal referents. *J. Exp. Psychol. Hum. Percept. Perform.* **2016**, *42*, 78. [CrossRef] [PubMed]
61. Sousa, M.; dos Anjos, R.K.; Mendes, D.; Billinghurst, M.; Jorge, J. WARPING DEIXIS: Distorting Gestures to Enhance Collaboration. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019; ACM: New York, NY, USA, 2019; p. 608.
62. Bangerter, A.; Oppenheimer, D.M. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture* **2006**, *6*, 85–102. [CrossRef]
63. Wong, N.; Gutwin, C. Where Are You Pointing? The Accuracy of Deictic Pointing in CVEs. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, Atlanta, GA, USA, 10–15 April 2010; ACM: New York, NY, USA, 2010; pp. 1029–1038. [CrossRef]
64. Bailenson, J.N.; Blascovich, J.; Beall, A.C.; Loomis, J.M. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence* **2001**, *10*, 583–598. [CrossRef]
65. Andersen, P.A.; Leibowitz, K. The development and nature of the construct touch avoidance. *Environ. Psychol. Nonverbal Behav.* **1978**, *3*, 89–106. [CrossRef]